WITHIN-HOST MICROBIAL INTERACTIONS AND PLANT PARASITES: FROM PAIRWISE INTERACTIONS TO THE MICROBIOME

Kayleigh Rose O'Keeffe

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology.

Chapel Hill
2019

Approved by:

Charles Mitchell

Ignazio Carbone

Corbin Jones

Elizabeth Shank

James Umbanhowar

ABSTRACT

KAYLEIGH ROSE O'KEEFFE: Within-host microbial interactions and plant parasites: from pairwise interactions to the microbiome
(Under the direction of Charles E. Mitchell)

Multicellular organisms are often host to a diverse community of mutualistic, commensal, and parasitic microbes, referred to collectively as the microbiome. The microbial community surrounding a parasite shapes both that parasite's immediate phenotype and its evolutionary potential. This dissertation investigates this by focusing on how within-host interactions relate to disease at multiple levels, from interrogations of pairwise interactions to the microbiome. In this work, I used computational methods and lab- and field-based studies to examine the interactions between fungal species that coinfect plant host individuals and co-occur in plant host populations.

At the individual level, I explored how interactions among fungal symbionts in the same host leaf alter parasite growth and host responses. Within host leaves, the growth of a parasite, *R. solani*, was influenced by coinfection with another parasite. This effect was further mediated by a third, asymptomatic fungal symbiont, highlighting the potential for higher-order interactions occurring among symbionts within a shared host. Multiple infection by these three fungal symbionts also impacted leaf survival and host biomass. I also simulated dual RNA-seq datasets, simultaneous gene expression sequencing of a pathogen and host during infection, to explore the challenges of investigating the molecular

mechanisms underlying host-parasite interactions in non-model systems and proposed a workflow to follow to interrogate such systems.

Scaling up to the population level, I explored how interactions between parasite, *R. solani*, and a mutualist affect parasite spread through a host population. While host individuals with the mutualist had higher biomass, populations in which the mutualist was present experienced higher peak parasite prevalence than populations in which the mutualist was absent. I further explored these fungal symbionts under field conditions, by investigating how the diversity and composition of the fungal community of leaves associate with disease symptoms. Symptoms of parasite, *R. solani*, were associated with lower fungal richness and diversity, as well as distinct fungal composition, compared to asymptomatic leaves or leaves symptomatic of other symptoms. Together, these results highlight the utility of multi-level approaches to disease ecology and evolution.

To Mike, for truly being my partner in this journey, and to my parents, for a lifetime of support and for instilling in me a love of learning

ACKNOWLEDGEMENTS

The completion of this dissertation would not have happened without the support of many mentors, collaborators, friends, and family. First, I would like to thank my committee members. Corbin Jones supported me as I explored a question that involved methods that were wholly out of my comfort zone and engaged with my research questions in ways that made me really consider how my research contributes to the field as a whole. Ignazio Carbone's curiosity and passion for research is contagious, and he was someone I could rely on for methodological support, as well as conversations that would increase my excitement over my questions. James Umbanhowar was always willing to meet with me to provide statistical and theoretical advice. As the other members of my committee are all PIs on the same grant, Beth Shank was an important addition to my committee to give me a different perspective on my research and help me figure out how to most effectively communicate my findings to a broad academic audience. Charles Mitchell has been the most amazing mentor. He gave me the space to develop my own questions, even when they were in directions outside of his own expertise. He makes his trainees a top priority, and while he gave me space to grow, he was accessible in every way to work through questions and results. He also has been so supportive of both my long-term career goals and personal priorities, for which I am incredibly grateful.

I would like to acknowledge my funding sources. Much of my research was funded by the NSF-USDA joint program in Ecology and Evolution of Infectious Diseases (USDA-

NIFA AFRI grant 2016-67013-25762). My research was also funded in part by Amherst College's John Woodruff Simpson fellowship. The UNC Graduate School Dissertation Merit Assistantship provided tuition support and a stipend for my first year of graduate school. UNC's Alma Holland Beers Scholarship provided summer funding which allowed me to perform full-time fieldwork in the summers of my first and second year of graduate school. The Triangle Center for Evolutionary Medicine provided tuition support and a stipend for Spring 2016. The NSF Graduate Research Fellowship Program provided three years of funding allowing me to focus entirely on my graduate studies. When not supported by fellowships, the Department of Biology provided a TA-ship to support me.

I have benefitted tremendously from my labmates. Fletcher Halliday, Rob Heckman, Miranda Welsh, and Julie Geyer provided important insight and feedback. They helped me develop my research questions, improve my writing, and get me re-excited about my research when I was in the weeds with fieldwork or statistics. Erin Mordecai and I only overlapped in the lab for a few months, but during that time, she helped me reflect on the questions I was considering pursuing, and since, has been someone I have been able to continue to discuss my work with and rely on for support. Fletcher Halliday has become a wonderful collaborator and friend, and I am so thankful that we were able to integrate our respective interests into a project that ultimately did not make it into this dissertation. I look forward to continuing to work with him in the future.

I have been incredibly lucky to work with two amazing undergraduates, each for a few years, during my time in graduate school. Anita Simha started in the lab right around the time that I joined the lab, and her interests made me think about my own work differently. Her work ultimately motivated the final experiment in a chapter of this dissertation. Brandon

Wheeler played an incredible role in the development and collection of data for the experiment described in a chapter of this dissertation, and I loved being able to mentor someone who is so passionate about plants and ecology.

My research benefitted from assistance from our incredible lab manager, Brooklynn Newberry, and seasonal field technicians. In particular, Charlie Muirhead, Storm Crews, and Julie Long helped set up the field experiment that was the focus of one of the chapters of this dissertation. Brooklynn Newberry has been a marvelous addition to the lab, and her organization and willingness to help whenever I needed has been invaluable.

During graduate school, I have had the opportunity to develop both my teaching and science communication skills, both of which shaped how I think about science and will also be essential to reach my long-term career goals. Mara Evans has served as a role model for the type of instructor I would like to be and, in the process, has become a good friend. The ComSciCon community, has given me role models and peers within science communication, and my experiences attending and organizing the Research Triangle workshop were an important part of my graduate school career.

I want to thank my family and friends. My parents, Eva and Kevin O'Keeffe, were always a phone call away whenever I needed to talk and have always been incredibly supportive of my education. Much of my success and happiness is because of their foundational support. My friends in the biology department, namely, Catie Alves, Catherine Chen, Sara Snell Taylor, Anaïs Monroy-Eklund, and Laura Mudge, became lunch companions, study buddies, writing partners, and CrossFit friends, and being able to rely on them to work through ideas or just generally provide support has been critical over the years. Jessie McGinty moved down to Chapel Hill with me from Boston, and she helped me

immensely with my transition to graduate school (and always had delicious baked goods to share)! Many of my closest friendships are with friends who live far away, and even with the distance, have been so important to me during my time in graduate school. Thank you especially to Fabiana Kreines, Kate Berry, and Catherine Schwartz.

I want to thank my therapist, Margaret Rhee, for helping me find balance in my life.

Finally, I want to thank my partner Mike. We had not been dating long before I packed up and moved from Boston to Chapel Hill. We spent my first four years of graduate school apart, but with many flights and even more Skype calls, you became my main source of support and my anchor. Your understanding and willingness not just to listen to me as I question myself, but provide an important voice to boost my confidence during times when I needed it, allowed me to push through and succeed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Species interactions are defined by the effects that two species have on one another and have been studied extensively in free-living species (Gause & Witt, 1935; Hairston, Smith, & Slobodkin, 1960; Holling, 1959). Those free-living species are often host to a diverse set of mutualistic, commensal, and parasitic microbes, referred to collectively as the microbiome. Within the microbiome, parasites are among the best-studied microbes due to their negative impacts on their hosts (Borer, Kinkel, May, & Seabloom, 2013). Infectious disease has historically been studied in a one host-one parasite framework, but the field is now appreciating the role and complexity of within-host microbial interactions. Within a shared host, parasites may interact by directly facilitating or interfering with each other, or indirectly, mainly via host resources or the host's immune system (Alizon, de Roode, & Michalakis, 2013). Parasites may also interact with non-parasitic microbes within a host (Adame-Álvarez, Mendiola-Soto, & Heil, 2014; May & Nelson, 2014).

Interspecific interactions are defined by the actions, traits and density of individuals of one species and the effect of those on the individuals and population of another species. They are also important density-dependent factors shaping population dynamics through effects on mortality, reproduction, population growth rate, and population density. It has therefore been critical to assess these interactions at different scales to gain insight into the ecological and evolutionary consequences of them. Similarly, it is critical to assess how interactions between microbes within a shared host affect parasites at varying scales (O'Keeffe, Carbone, Jones, & Mitchell, 2017).

A key question in infection biology is how do these interactions shape parasites in both the near and long term. Within-host microbial interactions can alter disease severity, transmission rates, and host mortality. Near-term responses of parasites to microbial interactions can be observed in changes in parasite growth rates and subsequent disease severity under coinfection (Pieter T J Johnson & Hoverman, 2012; Pinna et al., 2016). Sequencing advances have facilitated the investigation of the mechanisms underlying these interactions, though have been limited to model systems with developed genomic resources (A J Westermann, Gorski, & Vogel, 2012).

In contrast to these near-term responses of parasites within host individuals, long-term responses of parasites to microbial interactions have included impacts on parasite transmission rates across host individuals (Susi, Barrès, Vale, & Laine, 2015), spatial and temporal patterns of parasite abundances (Halliday, Umbanhowar, & Mitchell, 2017), and species co-occurrence (P. A. Clay, Dhir, Rudolf, & Duffy, 2018). Such long-term consequences of within-host microbial interactions have historically received less attention. As recent evidence suggests that coinfection can potentially lead to opposite outcomes at the host individual and population scales (Gorsich et al., 2018), it is critical that the consequences of within-host microbial interactions be studied and integrated across levels.

The microbial community surrounding a parasite shapes both that parasite's immediate phenotype and its evolutionary potential. This dissertation investigates this by focusing on within-host interactions on disease at multiple levels, specifically, the individual, population, and microbial community context. I performed four experiments. First, I computationally investigated potential limitations of applying sequencing advances to non-model host/parasite systems and proposed a workflow to overcome such limitations. Second,

I investigated the near-term responses of parasites to within-host microbial interactions within a host individual in a series of inoculation experiments. Third, I investigated the long-term responses of parasites to within-host microbial interactions by quantifying effects on parasite spread through a population in a field mesocosm experiment. Fourth, I examined the community context of this system by examining how fungal diversity associates with disease symptoms in a metabarcoding survey of field-collected samples. The final three chapters investigate the same disease system across levels.

CHAPTER SUMMARIES

In Chapter 2, I explored the challenges of investigating the molecular mechanisms underlying host-parasite interactions in non-model systems (i.e. study systems in which genomic resources are limited). Dual RNA-seq simultaneously profiles the transcriptomes of a host and pathogen during infection and may reveal the mechanisms underlying host-pathogen interactions. Dual RNA-seq is inherently a mixture of transcripts from at least two species (host and pathogen), so this mixture must be computationally sorted into host and pathogen components. Sorting relies on aligning reads to respective reference genomes, which may be unavailable for both species in non-model host-pathogen pairs. To investigate if using the genomic resources of species closely-related to a pathogen species of interest can facilitate the use of dual RNA-seq in non-model systems, I performed a simulated sequencing experiment in collaboration with Corbin Jones. Specifically, we simulated dual RNA-seq datasets of a host/pathogen system and assessed alignment accuracy of dual RNA-seq datasets and how accuracy varied with alignment algorithm and genetic distance between the species of focus and the reference species. We found that aligners that were able to map pathogen transcripts to the reference genome of a species closely related to the pathogen also

mismapped transcripts originating from the host to the pathogen's related reference genome. We ultimately provided evidence for a workflow for dual RNA-seq experiments with non-model systems that utilizes *de novo* assembly or concatenated reference information. This chapter was published in *Methods in Ecology and Evolution* (O'Keeffe and Jones 2019).

In Chapter 3, I focused on consequences of within-host microbial interactions on disease at the host individual-level. Microbes within a shared host can interact with each other, and when those microbes include parasites, such interactions can affect the diseases that they cause (O'Keeffe, Carbone, Jones, & Mitchell, 2017). Microbes may interact by directly facilitating or interfering with each other, or indirectly, mainly via host resources or the host's immune system (Tollenaere, Susi, & Laine, 2016). These interactions between coinfecting microbes can also range, from synergistic to antagonistic (Larimer et al. 2010, Morris et al. 2010, Bordes and Morand 2011). In collaboration with Anita Simha and Charles Mitchell, I investigated the effect of higher-order interactions, indirect interactions among species through effects on intermediate species (Wootton, 1994), on disease; specifically, I considered the interactions among two coinfecting parasites, *R. solani* and *C. cereale* and a systemic endophyte, *E. coenophiala* within a shared host. We found that *C. cereale* facilitated the growth of *R. solani,* and *E. coenophiala* had no direct effect on the growth of *R. solani*. However, *E. coenophiala*-infected plants that were co-inoculated with both parasites had significantly lower survival and biomass. Additionally, there was weak statistical support for an increase in the magnitude of the facilitative effect of *C. cereale* on *R. solani* in *E. coenophiala*-infected leaves. We demonstrated that as coinfections gain more attention, it may be important also consider higher-order interactions.

In Chapter 4, I focused on population-level consequences of within-host microbial interactions on disease. We hypothesized that infection with systemic grass endophyte, *E. coenophiala*, which has been shown to suppress growth of certain parasites, may also affect spread of parasites between hosts. To determine the effect of *E. coenophiala* on spread of fungal parasite, *R. solani,* through a host population, I set up a field mesocosm experiment in collaboration with Brandon Wheeler and Charles Mitchell. We set up contained populations of host grass species, tall fescue, and each population was comprised of entirely *E. coenophiala*-inoculated plants or entirely *E. coenophiala*-free plants. A focal plant was inoculated with parasite, *R. solani*. Following inoculation, we quantified parasite prevalence within each plot over time, as well as disease severity on tracked leaves over time. While parasite spread over time was unaffected by *E. coenophiala,* populations with *E. coenophiala* present had a higher peak prevalence of the parasite, *R. solani,* than populations that lacked *E. coenophiala*. Within plants, *E. coenophiala* conferred a benefit in growth, as plants that had *E. coenophiala* had significantly higher aboveground biomass. While biomass did significantly contribute to models of the peak parasite prevalence, *E. coenophiala* affected peak parasite prevalence beyond the effect of biomass. Additionally, individuals with *E. coenophiala* had slightly higher severity of parasite symptoms over time. Our results suggest that while *E. coenophiala* is a mutualist of tall fescue, it may not be a defensive mutualist in relation to *R. solani*.

In Chapter 5, I explored how within-host fungal communities associate with disease. Historically difficult to interrogate – largely because work was limited to identifying microbial species by symptoms or by using culture-based methods – the composition, dynamics, and functional biology of the microbiome are now being revealed by next-

generation sequencing technologies (Consortium et al., 2012; Lundberg et al., 2012). The microbiome may interact with parasites by competing for resources, releasing antimicrobial compounds, or altering a host immune response. Such interactions can lead to consequences for host health, making the host more susceptible to, tolerant of, or resistant to a parasite. At the same time, the microbiome is dynamic and the introduction of a parasite can lead to a disturbance in its diversity and composition. To determine how the within-host fungal community of a host plant associates with symptoms of three co-occurring fungal parasites, I conducted a high-throughput sequencing survey of the fungal community of a host population in collaboration with Fletcher Halliday, Corbin Jones, Ignazio Carbone, and Charles Mitchell. Barcoded amplicon sequencing of the fungal ITS region revealed that leaf segments that were symptomatic of *R. solani* had significantly lower fungal diversity and unique fungal composition compared to segments that were asymptomatic or symptomatic of other parasites. We put forward the hypothesis that *R. solani*, which kills host cells as it extracts resources, may act as a niche modifier, affecting its host environment, specifically creating necrotic tissue, and affecting how other microbes can colonize that tissue.

CHAPTER 2-CHALLENGES AND SOLUTIONS FOR ANALYSING DUAL RNA-SEQ DATA FOR NON-MODEL HOST-PATHOGEN SYSTEMS

Introduction

Viruses, bacteria and fungi can invade and parasitize eukaryotic host cells. Hosts may respond to infection by upregulating defense pathways. Pathogens, in turn, evade these host immune responses as they infect and cause disease. As this process unfolds and each organism responds to the other, gene expression changes in both the host and the pathogen (Kawahara et al., 2012). Yet, despite the importance of host–pathogen interactions, the genetic mechanisms underlying host–pathogen interactions during infection remain poorly understood. (A J Westermann et al., 2012).

Gene expression studies have been revolutionized by RNA-seq (Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008). In infection biology, RNA-seq can involve sequencing RNA extracted from pathogen-infected host tissue, an approach Westermann et al. (2012) coined as dual RNA-seq. Dual RNA-seq allows gene expression profiles of the host and pathogen to be characterized simultaneously during their interaction. For example, Teixeira et al. (2014) illustrate that the fungal parasite *Monoliophthora perniciosa* orchestrates the changes in the metabolism of the cacao plant *Theobroma cacao* that increase the availability of nutrients before the pathogen ultimately kills the plant. Yet, despite the power of dual RNA-seq as a tool to identify the genetic mechanisms underlying host–parasite interactions, a number of complications make dual RNA-seq difficult to adapt to non-model systems.

Dual RNA-seq data are inherently a mixture of transcripts from both host and pathogen. Given this mixture, the host and pathogen reads need to be sorted from each other. This sorting is typically done by an alignment algorithm that maps the reads to the two reference genomes, that of the host and that of the pathogen. To increase the accuracy of sorting and limit the potential for reads to mismap to the wrong reference genome (i.e. pathogen reads mapped to the host genome and vice versa), dual RNA-seq studies have employed a variety of analytical approaches such as discarding reads that map to both the host and pathogen reference genomes (e.g. Alexander J Westermann et al., 2016) or concatenating host and pathogen reference genomes into a composite for genome alignment (Aprianto, Slager, Holsappel, & Veening, 2016). However, both methods of read sorting rely on reference genomes, limiting their application to systems in which the host and/or pathogen species have developed genomic resources. As model species with sequenced and annotated reference genomes comprise a small fraction of total species (**https://www.ncbi.nlm.nih.gov/genome**), many host–pathogen systems are potentially excluded from dual RNA-seq analysis.

For organisms without complete reference genomes, approaches to analyzing RNA-seq data include assembling reads de novo (i.e. assembling many small reads together into fewer longer fragments, Grabherr et al., 2011) and mapping reads to reference genomes of related species (Ekblom & Galindo, 2010). As dual RNA-seq contains a mixture of transcripts from the host and pathogen species, the implications of the choice of one of these methods are complex, as accurate quantification levels of gene expression require reads to be separated correctly, and can influence the accuracy of the biological insight gleaned from the data. Reference-based methods, especially when using the reference genomes of a related

species, could result in reads not mapping if the genetic distance between the target species and the reference species is too high. Alternatively, aligners that allow more mismatches between reads and the reference genome could mismap reads from the wrong species, which could lead to spurious inference of how gene expression is affected by infection. De novo assembly can alleviate some of these problems, by creating larger fragments with which to map, but de novo assembly may result in reads from the two species assembling into longer fragments, which would lead to inaccuracies in subsequent mapping. Furthermore, dual RNA-seq datasets vary in the proportion of pathogen reads to host reads in the datasets, depending on the system of interest (Baddal et al., 2015; Choi, Aliota, Mayhew, Erickson, & Christensen, 2014; Hayden et al., 2014), and this variation may affect the accuracy and biological interpretation of various analytical methods.

Despite the variety of analytical approaches and their potential influence on the interpretation of dual RNA-seq data, the accuracy of these methods has not been systematically assessed. Here, we determined if and how dual RNA-seq can be utilized in non-model host–pathogen systems in which genomic resources are limited. Specifically, we investigated if using the genomic resources of species closely related to the pathogen species of interest can facilitate the use of dual RNA-seq in non-model systems. To do this, we simulated dual RNA-seq datasets. Simulations allowed us to manipulate dataset characteristics, like the proportion of pathogen to host reads, as well as facilitated downstream assessments of the accuracy of various analytical approaches. To investigate if using the genomic resources of a species closely related to a species of interest to analyze dual RNA-seq is appropriate, we assessed how genetic divergence between the genome of the pathogen of interest (from here on referred to as the target genome) and the reference

genome affected mapping accuracy. Additionally, we assessed how this effect was mediated by different alignment methods and the fraction of pathogen reads in the sample (Figure 2.1). We explored four different approaches: 1. aligning raw reads to the reference genome of the pathogen or a related species, 2. aligning reads to the host genome first and then mapping those reads that did not map to the host genome to the reference genome of the pathogen or a related species (referred to as the sequential approach), 3. aligning reads to a composite genome comprised of the genomes of the host species and pathogen or a related species (referred to as the place-to-go approach), and 4. assembling reads de novo prior to aligning (referred to as the assembly approach). This assessment provides guidance as to how to approach dual RNA-seq when studying organisms that do not have fully sequenced genomes.

Methods

Study system

Simulating RNA-seq relies on a reference genome and annotation file as inputs; therefore, we model dual RNA-seq using well-characterized genomes for both host and pathogen. Additionally, to investigate the effect of genetic distance between the pathogen of interest and the reference genome used for aligning reads, we needed to model a pathogen species for which closely related sister species were also fully sequenced. First, we used *Arabidopsis thaliana* and *Schizosaccharomyces octosporus* to represent host and pathogen species, respectively. While this is not a naturally occurring host–pathogen system, or symbiosis for that matter, as model organisms, these species have fully sequenced and well-annotated genomes, which were ideal for our approach. Additionally, most species within the *Schizosaccharomyces* genus have sequenced and annotated genomes (Table A2.1). Using the genomes of the other *Schizosaccharomyces* species as references for read mapping

allowed us to assess if dual RNA-seq data could be analyzed by using the genomic information of a related species as a reference when studying a species without a sequenced genome. Therefore, *A. thaliana* and *S. octosporus* allowed us to quantify how sensitive (or robust) different potential analysis methods were to increasing genetic distances between the focal pathogen and the reference genome. *A. thaliana* will be referred to as the host, and *S. octosporus* will be referred to as the pathogen.

RNA-seq simulations

Flux Simulator was used to generate simulated RNA-seq data (Griebel et al., 2012). Flux Simulator produced sequencing reads from a reference genome according to annotated transcripts. RNA-seq data were simulated separately for the host and pathogen. For each, four datasets were simulated for a factorial combination of two read lengths (76-bp or 150-bp) and configuration (single end or paired end). Each dataset included 10 million reads, similar to the simulations used in Baruzzo et al., 2016. All other simulation parameters were run as default. Simulated datasets were output as FASTQ files. Reads within each dataset were labelled with unique species-identifying tags (either 'HOST' or 'PATH') to facilitate downstream assessments of alignments.

To create dual RNA-seq datasets, we randomly selected reads from complementing datasets (same read length and configuration) of the host and pathogen and mixed them together. For each of the four sets of sequencing parameters, we created 12 10-million read datasets that ranged from 1 to 100% pathogen reads. As dual RNA-seq is sequenced from RNA extracted from pathogen-infected host tissue, typical datasets are comprised of a very low percentage of pathogen reads. The analysis of datasets with higher percentages of pathogen reads was conducted to show when and how patterns changed across the range of

the percentages of pathogen reads. Some systems investigate simultaneous gene expression of host and pathogens by mechanically separating cells of each species prior to sequencing (Ellison, DiRenzo, McDonald, Lips, & Zamudio, 2017). Datasets with higher percentages of pathogen reads that still include host reads could represent sequencing from RNA extracted after imperfect cell sorting.

Dual RNA-seq analysis approaches

For each reference-based approach, the reference genomes and annotations for *S. octosporus* (the target pathogen)*, S. cryophilus, S. japonicus and S. pombe* were used, downloaded in April 2018 from Fungi Ensembl ((Rhind et al., 2011), Table A2.1). As *S. octosporus* was simulated as the pathogen within the generated dual RNA-seq datasets, the genomic resources for the other species within the *Schizosaccharomyces* genus facilitated the investigation of how different levels of evolutionary distance between the genome of the target species and reference genome affect mapping accuracy. To generate a reference transcriptome, the BEDTools 'getfasta' utility version 2.25.0 was used to extract the transcript sequences from each of these downloaded genomes as specified by coordinates in the complementing annotation files (Quinlan & Hall, 2010). These transcript sequences were used as reference transcriptomes for alignment of reads from each dual RNA-seq dataset.

Read mapping was conducted with four different aligners: TopHat2, STAR, MapSplice2 and NextGenMap. TopHat2 (version 2.1.1) and STAR (version 2.5.1b) (Dobin et al., 2013; Trapnell, Pachter, & Salzberg, 2009) are both splice-aware aligners. TopHat2 relies on a Burrows–Wheeler transform and FM-index to search for matches between a reference genome and RNA-seq reads. STAR, which uses a seed and anchor approach based

on a Maximal Mappable Prefix, is more robust to non-continuous reads and some mismatches. Default parameter settings were used for both methods.

In addition to these two splice-aware aligners, a de novo aligner, MapSplice2 (version 2.2.1) was used to map reads to a reference genome (Wang et al., 2010). MapSplice2 detects splice junctions without any dependence on splice site features (an annotation file). We also mapped reads with an unspliced aligner, NextGenMap (version 0.4.12) to map reads from each simulated dataset to each reference transcriptome (H. Li & Durbin, 2010; Sedlazeck, Rescheneder, & von Haeseler, 2013). The hash-based variable mismatch threshold algorithm of NextGenMap maximizes its ability to utilize divergent reads. Default parameter settings were used.

Using the genomic resources and aligners discussed above, we processed reads in four different approaches to investigate the effectiveness and accuracy of analytical methods for dual RNA-seq data, (Workflow in Figure 2.1):

1. Raw read mapping

First, we investigated the accuracy of mapping the raw sequencing reads that were comprised of both host and pathogen reads. We conducted alignments with the reference genome of the target pathogen species, *S. octosporus*, and those of species closely related to the target species. Each of the four alignment algorithms discussed above was utilized.

2. Sequential mapping approach: map to host genome first

While the decision of mapping first to the host or pathogen is somewhat arbitrary, we believed that mapping to the host first would provide insight into the potential biases of dual RNA-seq and the impact of unintentional sequencing of a pathogen along with the host (i.e. unwittingly sequencing an infected host). Thus, we first mapped simulated dual RNA-seq

datasets to the host genome, then took the reads left unmapped and mapped them to the genome of the target species and those of closely related species. Reads from each simulated dataset were mapped to the host genome using TopHat2 under default parameters. Following alignments, the output BAM files containing unmapped reads were converted to FASTA files using the 'bam2fq' function within Samtools version 1.3.1 (H. Li et al., 2009). These reads that did not map to the host genome were then mapped to each of the *Schizosaccharomyces* genomes with NextGenMap and STAR.

3. Place-to-go approach: mapping to concatenated genome

We further investigated potential alignment methods when a host genome is available by mapping reads to concatenated genomes of the host genome and either the target pathogen genome or the genome of a species closely related to the target pathogen species. First, we investigated mapping accuracy with the composite genome of the host *A. thaliana* and of the target pathogen species *S. octosporus*, the two species used to simulate the dual RNA-seq datasets. Second, to assess the effectiveness and accuracy of this method when only the genomes of species closely related to the target pathogen species are available, we also created composite genomes of *A. thaliana* and each of the three other *Schizosaccharomyces* genomes. Read mapping was conducted with the four different aligners as described above.

4. Assembly approach: de novo assembly

We investigated whether de novo assembly of reads prior to mapping affected the effectiveness and accuracy of alignments. Prior to mapping reads, Trinity (Haas et al., 2013, version 2.2.0) was used for *de novo* assembly. Default parameters were used. To determine which reads comprised each contig, Bowtie2 (version 2.3.4.1), was used to align reads back to the assembled contigs (Langmead & Salzberg, 2012). If only pathogen reads mapped to a

contig, the contig was tagged ‘pathogen’. If only host reads mapped to a contig, the contig was tagged ‘host’. If both pathogen and host reads mapped to a contig, the contig was tagged ‘undetermined’. After Trinity assembly and tagging, contigs were mapped to each of the Schizosaccharomyces genomes with NextGenMap and STARlong under default settings.

Evaluation of alignments

SAM/BAM conversions, sorting and indexing were performed with SAMtools version 1.3.1 and Picard version 2.2.4 (H. Li et al., 2009). For each alignment, the number of mapped and unmapped reads originating from *A. thaliana* and *S. octosporus* was counted by parsing BAM files for the previously added unique tags for each species.

To investigate how biological insight would be affected by alignment method, gene-wise counts were obtained with featureCounts (Liao, Smyth, & Shi, 2013) and differential gene expression analysis was performed following the instructions of the ‘DESeq2’ package (Love, Huber, & Anders, 2014) deposited in Bioconductor. Specifically, we quantified gene counts (reads overlapping exons as described in the annotation build) for alignments to the target pathogen genome. Gene expression was compared between alignments of the same sequencing dataset among the different aligner methods. These pairwise comparisons did not have replicates because our simulations were performed without stochastic sampling. To address that limitation, we used the rlog transformation function, which transforms the average of the genes across samples to a log2 scale, as well as accounts for genes for which the evidence for strong fold changes is weak due to low counts. This protocol does not produce *p*-values but provides a ranked list of genes by regularized fold changes.

Applications to other host–pathogen systems

To investigate if patterns observed with the above approaches held across other host–pathogen systems, we also simulated dual RNA-seq involving another fungal pathogen, *Candida albicans*, and a different host species, *Homo sapiens*. Additionally, we simulated a bacterial pathogen system with *Homo sapiens* and *Escherichia coli*. Many species within the *Candida* and *Escherichia* genera have reference genomes available. We utilized two sister species of each *C. dubliniensis* and *C. parapsilosis*, and *E. fergusonii* and *E. albertii* as reference species to evaluate impact on dual RNA-seq analytical methods (Table A2.2).

Results

Generation of simulated datasets

We simulated dual RNA-seq datasets to investigate if and how dual RNA-seq can be utilized in host–pathogen systems in which genomic resources are limited. Simulated datasets represented a factorial combination of read length (76 bp vs. 150 bp), sequencing configuration (paired or single ended), and 12 different ratios of host reads to pathogen reads. In total, 48 dual RNA-seq datasets, each with 10 million reads, were simulated. Similarly, we simulated dual RNA-seq datasets for *Homo sapiens* and *Candida albicans,* and *Homo sapiens* and *Escherichia coli*, also with varying ratios of pathogen reads to host reads. We will first discuss the main results from the 76-bp single-end *A. thaliana* and *S. octosporus* datasets relegating the extensions and ancillary results to the supplement.

Comparison of raw read mapping

We first assessed the accuracy of alignments of dual RNA-seq raw reads with different aligners, representing a cross section of alignment algorithms, when using the

correct target genome. When mapping raw reads to the target genome of the pathogen of interest, the four aligners had comparable mapping rates of reads originating from the pathogen (Figure 2.2A). TopHat2 and MapSplice2 aligned c. 88% of pathogen reads; STAR and NextGenMap each aligned over 99% of pathogen reads to the target reference. Mapping rate of pathogen reads was unaffected by the percentage of pathogen reads in the sequencing datasets. While STAR and NextGenMap achieved a high mapping rate of pathogen reads, both aligners also mismapped host reads to the genome of the target pathogen species (Figure 2.2B, A2.3). For STAR and NextGenMap alignments of datasets in which there were more host reads than pathogen reads, a common occurrence among real dual RNA-seq datasets, mismapped host reads comprised 25–98% of the total reads mapped. In sum, most pathogen reads from a dual RNA-seq experiment were aligned by common aligners, but the aligners that mapped the most pathogen reads to the source pathogen genome also incorrectly aligned the most host reads to the pathogen genome.

As many pathogen species do not have genomic resources, we investigated how each of the four aligners performed when mapping reads to the genome of a species closely related to the pathogen. TopHat2 and MapSplice2 performed poorly, mapping <0.5% of pathogen reads when mapping to any genome of a related species. STAR and NextGenMap were able to map pathogen reads when using the genomic information of a related species as a reference (Figure 2.2A). Mapping rates of pathogen reads remained unaffected by the percentage of pathogen reads in the sequencing datasets. When using STAR, c. 36% of pathogen reads mapped to the genome of *S. cryophilus*, the species most closely related to the target species, *S. octosporus*. 19% of pathogen reads mapped to the S. japonicus genome, and 3% mapped to the *S. pombe* genome (Figure 2a). When using NextGenMap, c. 60% of

pathogen reads mapped to the *S. cryophilus* transcriptome, 22% mapped to the *S. pombe* transcriptome and 28% mapped to the *S. japonicus* transcriptome (Figure 2.2A). For both STAR and NextGenMap, the percentage of pathogen reads mapped generally decreased as evolutionary distance between the target and reference genomes increased. Thus, there is a distinct bifurcation between aligners that can and cannot effectively map pathogen reads when only a related reference genome is available.

For STAR and NextGenMap, host reads mismapping to the incorrect genome increased as the evolutionary distance between the target pathogen species and the reference species increased, while TopHat2 and MapSplice2 only mapped a few host reads (maximum of 19 reads) to any of the *Schizosaccharomyces* reference genomes under any sequencing parameters (Figure 2.2B, A2.3). When aligning with STAR, c. 21% of host reads mapped to the *S. cryophilus* and *S. pombe* genomes. Although only c. 1% of host reads mapped to the *S. pombe* genome, only 3% of pathogen reads mapped to the same genome, so overall mapping rate was very low. When aligning with NextGenMap, c. 25% of host reads mapped to the *S. cryophilus* and *S. japonicus* transcriptomes, and c. 15% of host reads mapped to *S. pombe*. The effects of evolutionary distance-related mismapping is greatest in datasets in which the proportion of pathogen reads was low, as host reads comprise the majority of total reads mapped.

Gene counts and differential expression analysis

To investigate how biological insight would be affected by host read mismapping, we quantified gene-wise counts of the host reads that STAR mismapped to the target pathogen genome. The sequencing dataset that resulted in the highest number of host reads mismapped to the target pathogen genome in which the highest number of host reads mismapped was the

dataset with highest proportion of host reads relative to pathogen reads (specifically, 99% host reads/1% pathogen reads). The vast majority of the mismapped host reads derived from repetitive parts of the genome. As determined by featureCounts, only 1.6% of *S. octosporus* genes had more than 50 host reads mismap to them, and these genes varied in biological function (Table A2.3).

We compared two alignments of the same simulated dataset to the target genome of *S. octosporus*; one alignment was performed by mapping raw reads to the genome with TopHat2 (which did not mismap host reads) and the other was performed by mapping raw reads to the genome with STAR (which included host reads mismapping). 97.1% of genes that were expressed differently (above 0.1-fold change on the log2 scale) between alignments were overexpressed in the alignment in which host read mismapping occurred. This suggests that mismapping of reads to the wrong reference in dual RNA-seq can result in upward biases in estimates of gene expression. Thus, compared to an uninfected control, these upwardly biased genes would appear to be 'induced' by infection.

Alternative mapping strategies may reduce mapping problems

We considered three alternative approaches that could reduce poor mapping and mismapping of dual RNA-seq data. We first investigated approaches that would be possible if a host genome was available. We tried to filter out host reads by first aligning dual RNA-seq datasets to the host genome, and then mapping the reads left unmapped to the target pathogen genome (or those of related species). This 'sequential' approach decreased the amount of host read mismapping only slightly (Table 2.1). For most alignments, the percentage of total mapped reads that originated from the host only decreased by 1–3%.

We investigated a second approach in which reads were mapped to concatenated genomes of the host and either the target pathogen or that of a closely related species to the pathogen (a 'place-to-go' design). The place-to-go method resulted in alignments using STAR and NextGenMap having substantially fewer host reads mismapping to the genome of the pathogen or that of a species closely related to the pathogen compared to the alignments to genomes excluding the host genome. Furthermore, both aligners retained their ability to map pathogen reads to the genomes of closely related species to the target pathogen (Figure 2.3, Figure A2.4). Therefore, the place-to-go method may overcome some of the limitations of mapping to a related reference in dual RNA-seq.

Finally, we investigated a third approach in which reads were first assembled de novo into longer fragments, then those fragments were mapped to the genomes of the pathogen and species closely related to the pathogen ('assembly' approach). The majority of assembled contigs were comprised entirely of host reads or entirely of pathogen reads (Assembly metrics in Figure A2.1). A small fraction of contigs were chimeras—that is, a mix of host and pathogen reads (labelled as 'undetermined'). Alignments of these assembled transcripts to each of the reference genomes resulted in a substantial decrease in host read mismapping while preserving the ability to map pathogen contigs (Figure 2.4, Figure A2.5). Across all reference species and proportions of pathogen reads in the original datasets, <1% of contigs comprised of host reads mapped with NextGenMap. Furthermore, c. 99% of contigs comprised of pathogen reads mapped to the target transcriptome of *S. octosporus*. Approximately 79% of contigs comprised of pathogen reads mapped to *S. cryophilus*, 29% of contigs comprised of pathogen reads mapped to *S. pombe* and 15% of contigs comprised of pathogen reads mapped to *S. japonicus*. Although some of the contigs that were unable to be

identified as comprised of pathogen or host reads mapped, they comprised a minority of the total number of mapped contigs, with a maximum of 2.5%. With such a reduction in host read mismapping, the majority of all mapped contigs originated from the pathogen. Given that a good de novo assembly is possible, the assembly approach also clearly reduced mismapping.

Effect of sequencing parameters

To investigate the effect of sequencing parameters—the size of sequencing read, paired-end vs. single-end reads—on the above approaches, we simulated dual RNA-seq datasets of the same system with a longer read length (150 bp) and with paired-end sequencing (Figure A2.11–A2.16). While the same patterns largely held—that raw read mapping resulted in host reads mismapping when aligning with STAR and NextGenMap and mapping reads to a concatenated genome or assembling reads de novo prior to mapping substantially reduced host read mismapping—there were some differences among the layouts. Specifically, longer read lengths not only resulted in overall lower host read mismapping rates (which is consistent with the results of mapping assembled reads) but also resulted in lower pathogen read mapping rate, especially when mapping to the genomes of species closely related to the pathogen. Additionally, assembling 76 paired-end reads de novo prior to mapping to the *Schizosaccharomyces* genomes resulted in more reads that could not be identified as from the host or pathogen (potentially chimeras) comprising the group of mapped reads. As expected, longer paired-end reads generally performed better than other configurations.

Similarities and contrasts in other taxa

The simulations above focused on a fungal 'parasite' infecting a plant host as we believe that this could be a particularly problematic scenario, as plants and fungi are both eukaryotes, while other host–pathogen systems involve more diverged species. To investigate if and how the patterns observed above extend to other systems, we simulated two more sets of dual RNA-seq datasets. We simulated datasets across the same range of proportion of pathogen reads for another host–fungal pathogen system, *Homo sapiens* and *Candida albicans*, as well as a bacterial pathogen system, *Homo sapiens* and *Escherichia coli*. While alignments of the Human-*Candida* raw reads to the genomes of the target species, *C. albicans*, and two closely related species, *C. dublinensis* and *C. parapsilosis,* did result in a comparable level of host read mismapping to the above analyses (Figure A2.6), alignments of Human-*E. coli* raw reads had minimal if any host read mismapping (Figure A2.10). We conducted the same three approaches described above to minimize host read mismapping with the Human-*Candida* alignments, and we observed the same results as described above—that mapping to concatenated genomes of the host and closely related species, as well as de novo assembly prior to aligning, substantially reduce host read mismapping while retaining the ability to map pathogen reads (Figure A2.7–9). Therefore, the mismapping problems in dual RNA-seq held, to varying degrees, across the systems we investigated.

Discussion

Understanding the genetic mechanisms of host–pathogen interactions may be improved using dual RNA-seq (A J Westermann et al., 2012), but several limitations have contributed to the underutilization of this approach. Dual RNA-seq is inherently a mixture of host and pathogen reads that need to be parsed prior to analyses. This parsing relies on

mapping reads to the genomes of each organism. Consequently, it was previously unknown whether and how dual RNA-seq could be applied to non-model host–pathogen systems, in which there are limited or no genomic resources. Our analyses of simulated sequencing identified as problematic several approaches that might be encountered by researchers applying dual RNA-seq to non-model host–pathogen systems. However, our systematic comparison of analytical approaches also revealed a workflow that can be used to identify the genetic mechanisms of host–parasite interactions for non-model organisms.

For non-model organisms, traditional approaches to analyzing RNA-seq data include mapping reads to reference genomes of related species (Benjamin, Nichols, Burke, Ginsburg, & Lucas, 2014). Depending on the software for aligning sequencing reads, we found that using genomic resources of a closely related pathogen can result in one of two error modes. We found that aligners like TopHat2 and MapSplice2 are too restrictive with allowed mismatches by default, resulting in pathogen reads failing to map to genomes of the closely related species. In contrast, aligners such as NextGenMap and STAR are too lenient, allowing for too many mismatches by default and resulting in the mismapping of host reads to the genome of the species closely related to the pathogen. This was consistent when investigating simulated dual RNA-seq datasets of a plant (*Arabidopsis thaliana*) and fungus (*Schizosaccharomyces octosporus*) as well as simulated dual RNA-seq datasets of human and fungal pathogen *Candida albicans*. In contrast, host read mismapping was substantially reduced for a simulated dataset of human and bacterial pathogen, *Escherichia coli*, suggesting that this mapping inaccuracy may be a particular concern for studies of fungal pathogens.

The difference in the performance of the alignment tools likely reflects the classic trade-off between precision vs. sensitivity in the underlying algorithms. TopHat2 underutilizes the dual RNA-seq data as it relies on a Burrows–Wheeler transform and FM-index to quickly search for matches between the reference genome and the RNA-seq reads. This emphasis on fast and exact (or nearly so) match alignment struggles to map reads that have divergent bases. MapSplice2, a splice-aware aligner, applies a metric based on Shannon maximum entropy as applied to a weighted de Bruijn graph. This approach can detect splice junctions without any dependence on splice site features—potentially a critical feature when applying dual RNA-seq to poorly annotated genomes or closely related species. However, the base alignment approach relies on Bowtie algorithm and as a result suffers from the same limitations as TopHat2. In contrast, the hash-based variable mismatch threshold algorithm of NextGenMap maximizes its ability to utilize divergent reads but makes more erroneous assignments. STAR, which uses a seed and anchor approach based on a Maximal Mappable Prefix, is robust to non-continuous reads and some mismatches. It performed almost as well as NextGenMap in terms of data utilization, but again suffered from imprecision in the form of host reads mismapping to the wrong genome.

Host reads mismapping to the genome of the pathogen or the genome of a species closely related to the pathogen can have severe implications for the characterization of the gene expression profile of the pathogen during the infection process. Differential gene expression analyses between alignments of the same dual RNA-seq dataset that were produced by mapping raw reads to the target genome with TopHat2 (in which host reads did not mismap) and STAR (in which host reads did mismap) indicated that the alignments produced with STAR had overall higher levels of overexpression than that produced by

TopHat2. This highlights that biological insight gained from dual RNA-seq data can be inaccurate if certain steps are not taken. As real sequencing is unable to definitively identify the species origin of transcripts, neglecting to take measures to avoid host reads mismapping would result in inaccurate genetic mechanisms implicated in the infection process.

From our results, we propose a workflow that should be followed to determine the best approaches to extending the use of dual RNA-seq to a wider array of systems, including non-model systems (specifically, eukaryotic host–fungal pathogen systems) in which genomic resources are available for the species closely related to the pathogen of interest (Figure 2.5).

- If a host genome is available, concatenating the genome of the host with the genome of species closely related to the pathogen of interest (place-to-go approach) results in more accurate alignments, in which host read mismapping is substantially reduced, with the aligners, STAR and NextGenMap.
- If a host genome is not available, assembling reads de novo, prior to aligning with NextGenMap and STAR decreased host read mismapping while retaining the ability to map pathogen reads. As expected, de novo assemblies do exclude some reads, which would consequently not be quantified. The reads excluded, however, were rare and a subset of them was low-complexity sequences (Figure A2.2).

As NGS technologies and their analytical tools continue to become more affordable and accessible, it is important to critically assess how accurate genomic analyses with these tools are. While dual RNA-seq has been applied to many model disease systems, we remain woefully unaware of how the accuracy of mapping methods utilized to separate host and

pathogen reads affects dual RNA-seq studies. The methods we used here allowed us to assess the accuracy of alignment approaches of dual RNA-seq in non-model systems through simulated sequencing, but the biological truth of the origin of transcripts in real dual RNA-seq data would remain unknown and the issues we identified would lead to misinterpretations of the data. As infectious diseases are expected to increase in the coming years, it is critical that we investigate proper methods of analyses to ensure accurate insights are gained as systems are explored.

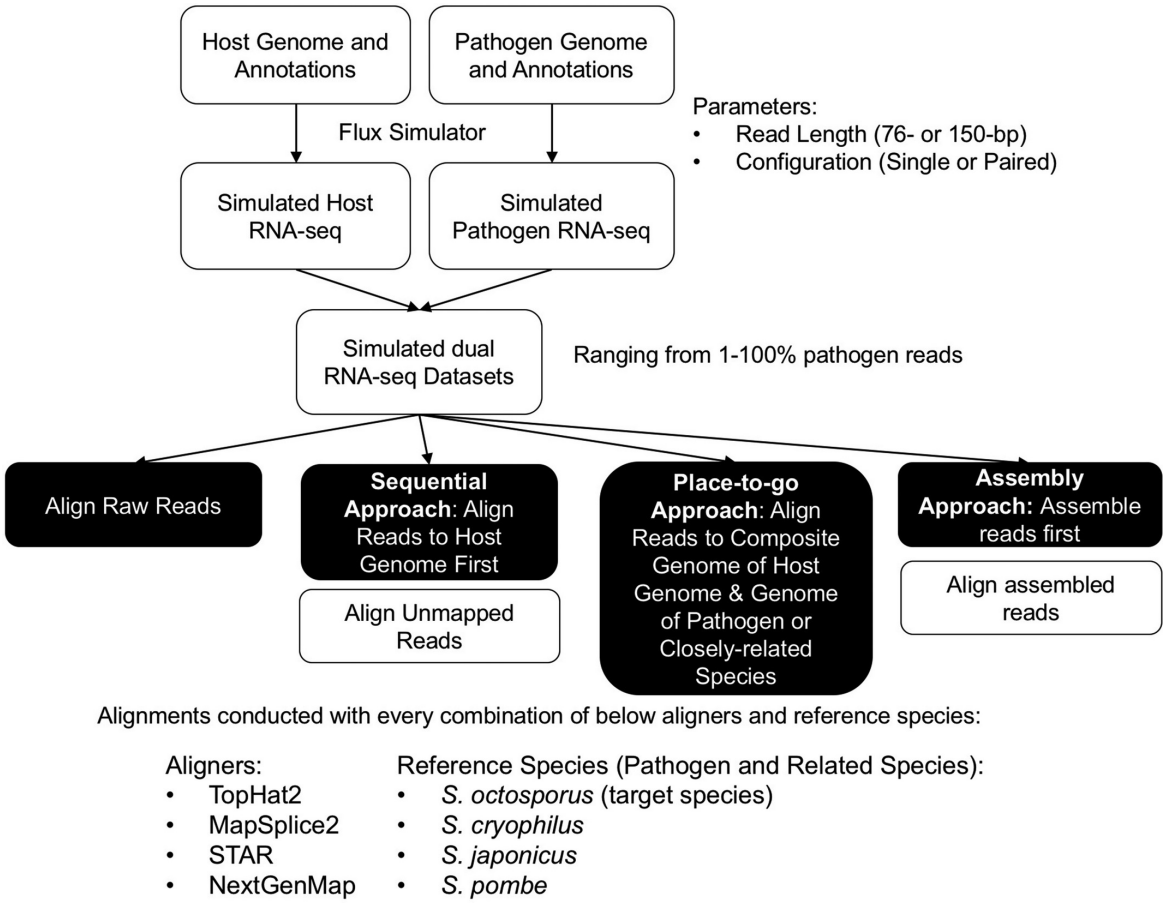


Figure 2.1: Dual RNA-seq Simulation Study Workflow. We outline our steps to investigate best approaches for analyzing dual RNA-seq datasets of non-model systems

Table 2.1: Results of Sequential Approach. Host mismapping rates (proportion of reads that mapped to pathogen genome or that of a closely-related species that originated from host) are shown for raw read approach in which sequencing reads were mapped without any prior steps and sequential approach, in which reads that did not map to an initial alignment to host genome were mapped. Results are shown for each factorial combination of reference species and the proportion of pathogen reads in the simulated dataset. Cells color indicates magnitude of host read mismapping rate: dark orange (>50% of total mapped reads are from host), intermediate orange (10-50%), light orange (<10%). When using NextGenMap, there are slight decreases when reads are first filtered by mapping to host genome first, but there is still a substantial amount of host mismapping with both methods. When using STAR, there is no effect of first filtering datasets by mapping to host genome first on host read mismapping.

Target species — *Increasing evolutionary distance from target* →

NextGenMap

Proportion of Pathogen Reads in Dataset	*S.octosporus*		*S. cryophilus*		*S. pombe*		*S. japonicus*	
	Raw Reads	After Host Mapping	Raw Reads	After Host Mapping	Raw Reads	After Host Mapping	Raw Reads	After Host Mapping
0.1	0.70	0.68	0.80	0.78	0.86	0.83	0.89	0.88
0.2	0.51	0.49	0.64	0.61	0.72	0.68	0.78	0.77
0.3	0.38	0.35	0.50	0.48	0.61	0.56	0.68	0.66
0.4	0.28	0.25	0.39	0.37	0.50	0.45	0.57	0.55
0.5	0.21	0.19	0.30	0.28	0.40	0.35	0.47	0.45
0.6	0.15	0.12	0.22	0.21	0.31	0.26	0.38	0.36
0.7	0.10	0.08	0.16	0.14	0.22	0.19	0.28	0.26
0.8	0.06	0.04	0.10	0.09	0.14	0.12	0.18	0.17
0.9	0.01	0.00	0.05	0.04	0.07	0.06	0.09	0.08
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

STAR

Proportion of Pathogen Reads in Dataset	*S.octosporus*		*S. cryophilus*		*S. pombe*		*S. japonicus*	
	Raw Reads	After Host Mapping	Raw Reads	After Host Mapping	Raw Reads	After Host Mapping	Raw Reads	After Host Mapping
0.1	0.67	0.67	0.84	0.84	0.9	0.9	0.83	0.83
0.2	0.47	0.47	0.7	0.7	0.8	0.81	0.69	0.69
0.3	0.34	0.34	0.57	0.57	0.71	0.71	0.57	0.57
0.4	0.25	0.25	0.46	0.46	0.61	0.61	0.46	0.46
0.5	0.18	0.18	0.37	0.37	0.51	0.51	0.36	0.36
0.6	0.13	0.13	0.28	0.28	0.41	0.41	0.27	0.27
0.7	0.09	0.09	0.2	0.2	0.31	0.31	0.19	0.19
0.8	0.05	0.05	0.13	0.13	0.21	0.21	0.12	0.12
0.9	0.02	0.02	0.06	0.06	0.1	0.1	0.06	0.06
1	0	0	0	0	0	0	0	0

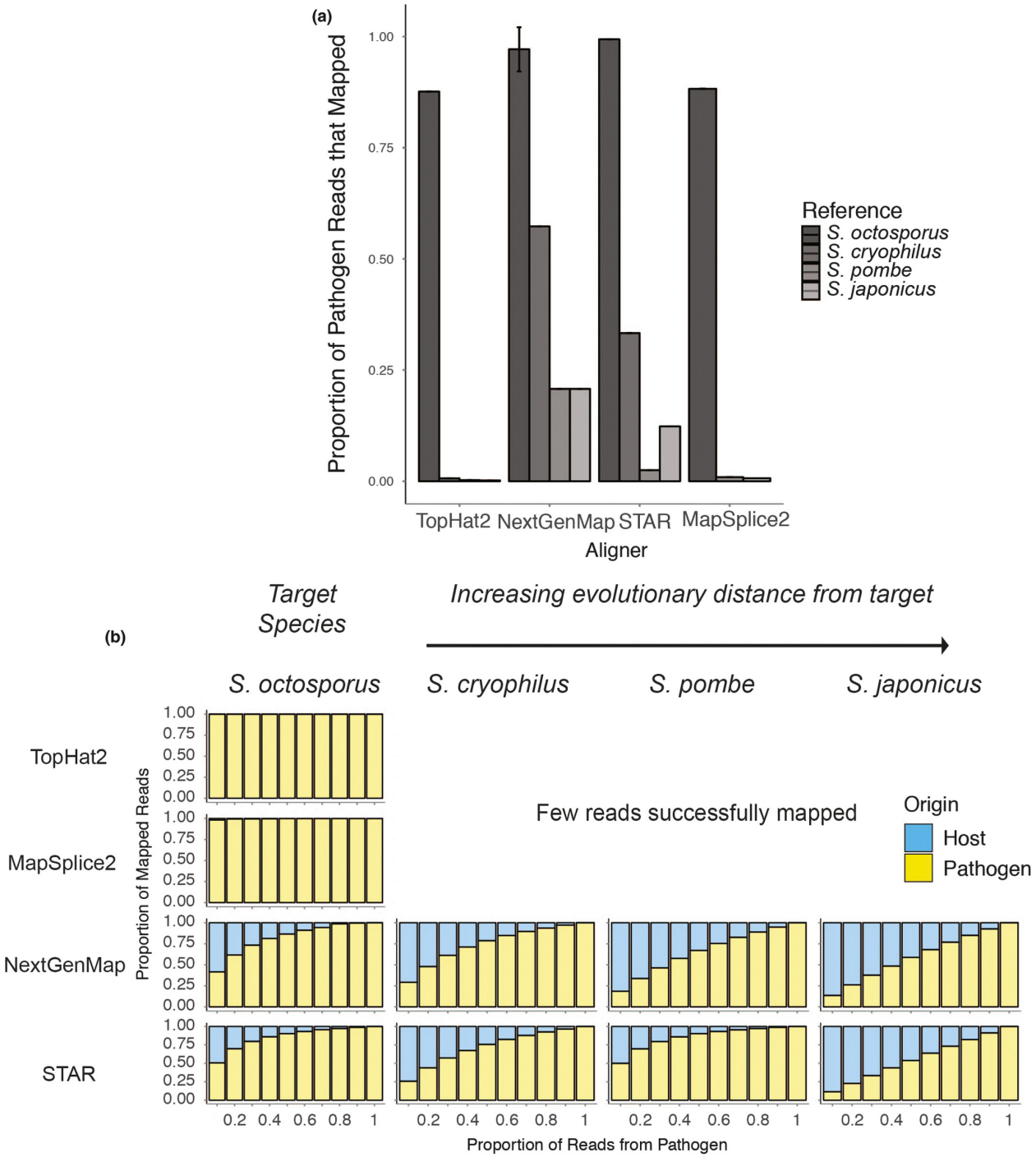


Figure 2.2: Comparison of utility and accuracy of several aligners for dual RNA-seq analysis. Factorial combination of four aligners and the four *Schizosaccharomyces* reference genomes. (a) Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to target pathogen species are shown in black, and greyscale gradient indicates evolutionary distance from the target. TopHat2 and MapSplice2 can only map pathogen reads when mapping to target pathogen genome and are unable to effectively map reads to genomes of related species. STAR and NextGenMap are able to map pathogen reads to reference genomes of

related species. (b) Origin of Mapped Reads. For each bar plot, blue reads are those that originated from host (*Arabidopsis thaliana*) and yellow reads are those that originated from pathogen (*Schizosaccharomyces octosporus*). TopHat2 and MapSplice2 can only effectively map pathogen reads when mapping to target pathogen genome, and are unable to effectively map reads to the reference genomes of related species. These plots are therefore excluded. STAR and NextGenMap are able to map pathogen reads to reference genomes of related species, but both aligners result in host reads mismapping

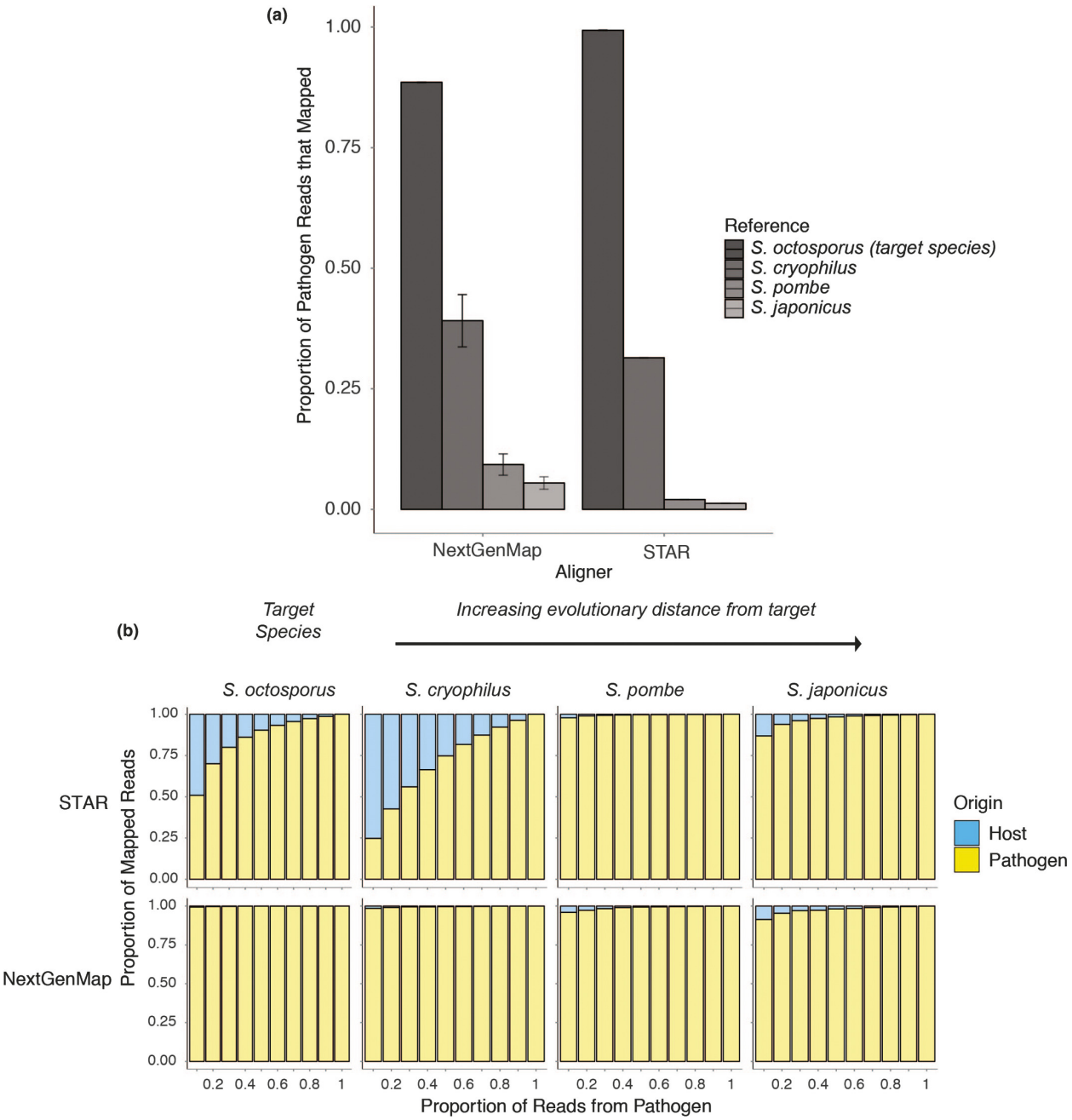


Figure 2.3: Concatenated genome mapping (Place-to-go approach) improve dual RNA-seq analysis. Results are shown for aligners STAR and NextGenMap and four composite genomes comprised of the *Arabidopsis thaliana* (host) genome and each of four *Schizosaccharomyces* reference genomes. (a) Bars indicate the proportion of pathogen reads that mapped to the part of the composite genome originating from the genome of the pathogen or closely related species to pathogen. Alignments to target genome of the pathogen species of interest are shown in black, and greyscale gradient indicates evolutionary distance from target. STAR is able to map almost all pathogen reads to target genome, and 27.7% of pathogen reads to the composite genome with genome of the most closely related genome to the target. NextGenMap is able to map c. 89% reads to target, and 39% reads to the composite genome with genome of most closely related genome to target. (b) Origin of reads that mapped to genome of

pathogen or closely related species. For each bar plot, blue reads are those that originated from host (*Arabidopsis thaliana*) and yellow reads are those that originated from target pathogen (*Schizosaccharomyces octosporus*). Bar plots represent the composition of reads that mapped to the component of each composite genome corresponding to either the target pathogen genome or the genome of a closely related species

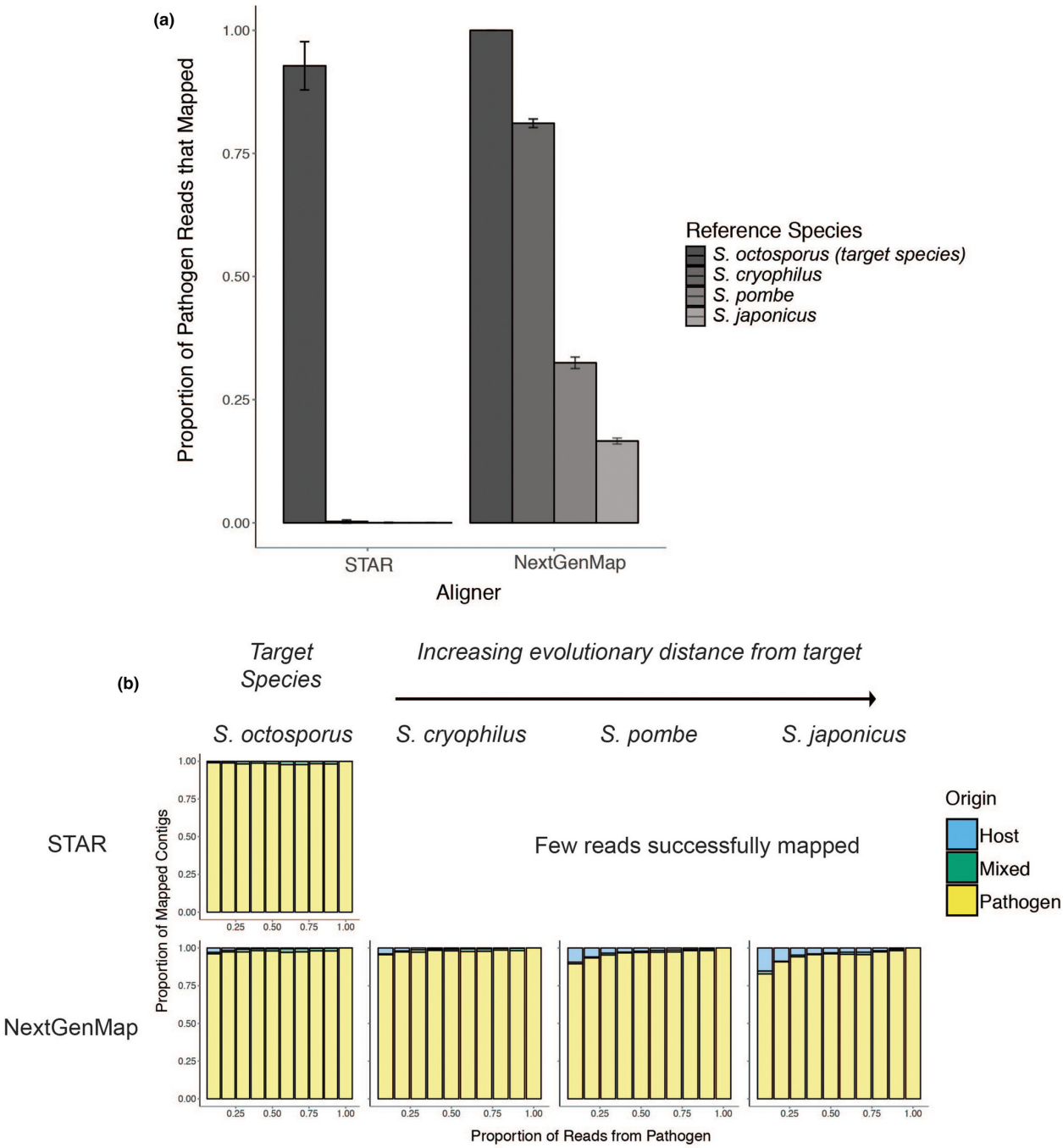


Figure 2.4: The Assembly Approach improves alignment of dual RNA-seq reads. Results are shown for aligners STAR and NextGenMap and four *Schizosaccharomyces* reference genomes. (a) Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to target pathogen species are shown in black, and greyscale gradient indicates evolutionary distance from target. STAR can only map pathogen contigs when mapping to reference genome of the pathogen and is unable to effectively map contigs to reference genomes of related species. NextGenMap retains its ability to map pathogen contigs to transcriptomes of related species. (b) Origin of assembled reads that mapped to genome of pathogen or closely related species. For each bar plot, blue contigs are those that originated from host (*Arabidopsis thaliana*) and yellow contigs are those that originated from pathogen (*Schizosaccharomyces*

octosporus). Contigs that were unable to be determined as comprised of host or pathogen reads are colored green. The bar plots represent composition of contigs that mapped to each reference genome. STAR was unable to align more than a few contigs for *S. cryophilus, S.pombe* and *S. japonicus,* so those plots have been excluded. Host mismapping was strongly reduced

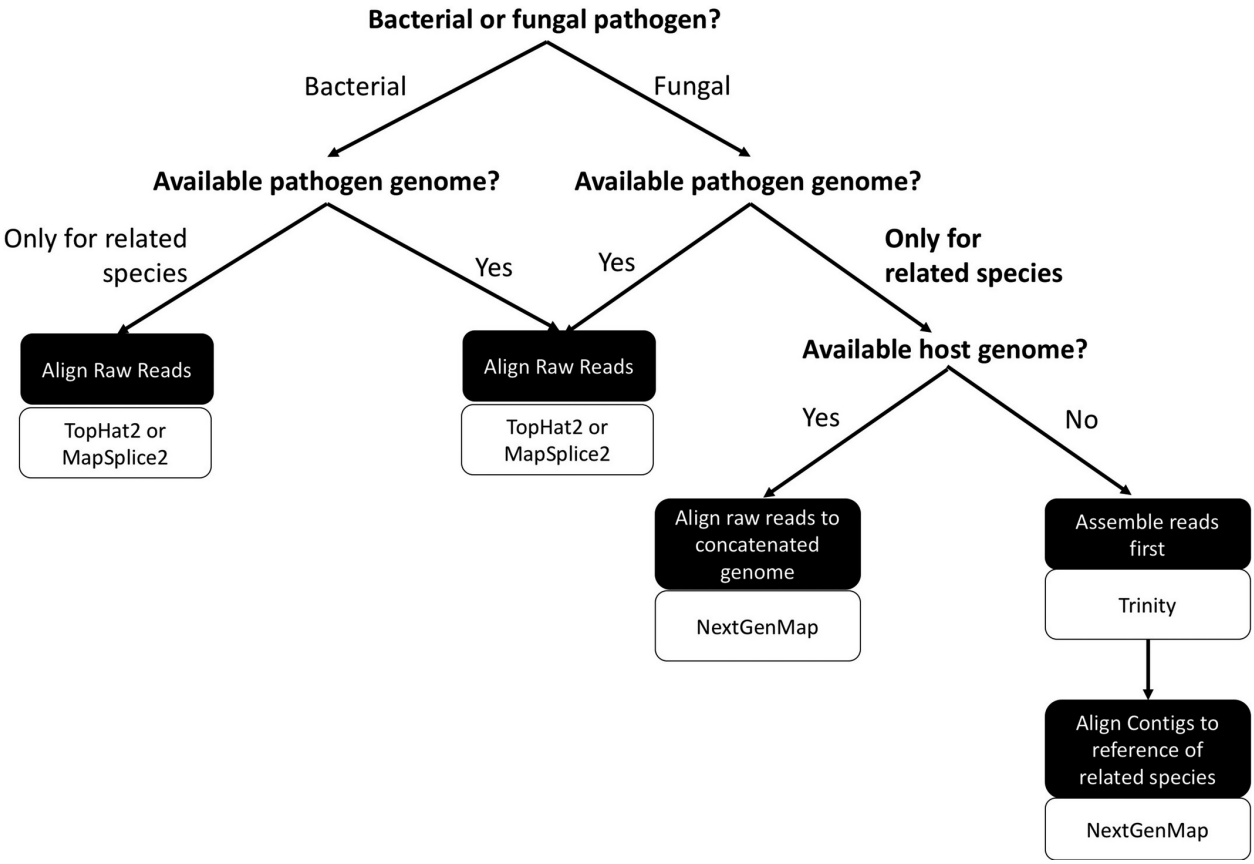


Figure 2.5: Suggested Workflow for dual RNA-seq Experiments of non-model host–pathogen systems

CHAPTER 3: HIGHER-ORDER INTERACTIONS BETWEEN PARASITES AND A SYSTEMIC GRASS ENDOPHYTE AFFECT PARASITE GROWTH AND HOST MORTALITY

Introduction

Multicellular organisms are host to a diversity of microbes, and these microbes—whether they be bacteria, fungi, or viruses—can range in their relationship with their hosts, from parasitic to mutualistic. While these terms are rooted in the interaction between a given microbe and its host, microbes within a shared host can also interact with each other (O'Keeffe et al., 2017). Microbes may interact directly, via chemical secretion or physical interaction, or indirectly, mainly via host resources or the host's immune system (Tollenaere, Susi, & Laine, 2016). These interactions between coinfecting microbes can range from antagonistic to synergistic. As the ubiquity of multiple infection has been realized, the interactions among coinfecting microbes and the effects of those interactions on each microbe, as well as the shared host, have gained a lot of attention (Larimer et al. 2010, Morris et al. 2010, Bordes and Morand 2011). Here, we focus on higher-order interactions, indirect interactions among species requiring an intermediary species in order to arise (Wootton, 1994); specifically, we consider the interactions among two coinfecting parasites and a hypothesized mutualist within a shared host.

Coinfection of a host with multiple parasites is common in nature (Cox FE, 2001) and can lead to changes in parasite growth rates, disease severity, host susceptibility, and infection length (Diuk-Wasser, Vannier, & Krause, 2016; Gibson et al., 2011; Halliday, Umbanhowar, & Mitchell, 2017, Seabloom et al. 2009, Johnson and Hoverman 2012). Further, the infection

sequence of coinfecting parasites on a host can determine the direction and magnitude of these effects of these interactions within hosts (Adame-Álvarez, Mendiola-Soto, & Heil, 2014; Hood, 2003; Lohr, Yin, & Wolinska, 2010; Pathirana, Meegaskumbura, & Rajakaruna, 2019). These interactions within host individuals can impact population-level processes, as parasite coinfections have major consequences for the spread of parasites in a population and epidemic dynamics (Halliday, Umbanhowar, et al., 2017; Telfer et al., 2010).

While interactions among parasites are increasingly studied in disease ecology, we have a more limited understanding of interactions of parasites with microbial mutualists (Hacquard and Schadt 2015, Johnson et al. 2015b). Microbial mutualists are ubiquitous among multicellular organisms and can confer a number of benefits to their hosts, including decreasing parasite infectivity (Haine, 2008). Microbial mutualists can limit parasite colonization and growth by altering nutrient availability or competing for metabolites within a shared host (Gerardo & Parker, 2014). They can also prime a host's immune system, which can facilitate a rapid response to invading parasites (Belkaid & Hand, 2014). Some microbial mutualists can affect parasite success more directly as well, by producing antimicrobial compounds that can reduce parasite growth and consequent disease severity (Clay 1990, Panaccione et al. 2014).

Multi-species interactions involving more than two species have been studied extensively within free-living species, showing that such interactions can be additive, non-additive or more complex (Darling & Côté, 2008; Sih, Englund, & Wooster, 1998). However, the study of within-host microbial interactions, especially those involving parasites, often focuses on the effects of pairwise combinations of species (but studies investigating more than pairwise interactions include Abbate et al., 2018; Marchetto & Power, 2018). Since the mechanisms by which two symbionts can interact are not mutually exclusive, the effects of coinfection may be influenced

by other co-infecting symbionts. For example, microbial mutualists can affect available host resources (Bronstein, 2001), which may modify the interaction between coinfecting parasite species that rely on these host resources (Wootton, 1994). Microbial mutualists could also be considered a shared enemy of coinfecting parasites, resulting in dynamics analogous to enemy-mediated apparent competition (Chaneton & Bonsall, 2000).

While the direct and indirect interactions that occur among species can shape ecological communities, empirical evidence of these interactions within microbial communities is limited. The detection of higher-order interactions can be difficult in ecological experiments because responses may contain both direct and indirect effects. Such approaches are further complicated in the context of microbial communities, in which the species can be difficult to detect and isolate. Here, we used a series of controlled inoculation experiments to disentangle higher-order interactions that act among fungal symbionts of a plant host. Specifically, we evaluated the interactions among a microbial mutualist and two coinfecting parasites of grass host, tall fescue.

Methods

Study System

We investigated the interactions among three fungal symbionts of tall fescue (*Lolium arundinaceum*): two parasites, *Colletotrichum cereale* and *Rhizoctonia solani*, and an endophyte, *Epichloë coenophiala*. *C. cereale* causes anthracnose in cool-season grasses. It is an obligate hemibiotroph; it has an initial biotrophic phase in which it extracts resources from living cells that is asymptomatic, then a second necrotrophic phase in which it kills host cells to extract resources, at which point symptoms appear. *R. solani* causes brown patch in grasses. In plants, it is a necrotroph, killing host cells to extract resources. It can also persist in soil as a saprobe.

Biotrophs often facilitate necrotrophs via immune-mediated crosstalk, and necrotrophs inhibit biotrophs via competition for host resources (Glazebrook, 2005). Given the hemibiotrophic feeding strategy of *C. cereale,* we would expect *C. cereale* in its initial biotrophic phase to facilitate *R. solani,* and for that to change when *C. cereale* shifts to its necrotrophic phase. Investigations of these hypothesized interactions in a field population of tall fescue showed *C. cereale* to consistently inhibit the growth of *R. solani* (Halliday, Umbanhowar, et al., 2017).

E. coenophiala is a vertically transmitted systemic fungal endophyte that is considered a mutualist under most ecological conditions (Kari Saikkonen, Young, Helander, & Schardl, 2016). *E. coenophiala* can increase its host's tolerance of and recovery from drought (Malinowski & Belesky, 2000) and produces bioactive alkaloids that can protect plants against herbivores (K. Saikkonen, Helander, Faeth, Schulthess, & Wilson, 1999). In the context of infectious diseases, *E. coenophiala* can facilitate or suppress infection by fungal parasites via production of toxins and changes in host immunity, depending on parasite feeding strategy (Potter 1980, 1982; Liu et al. 2006; Saikkonen et al. 2013). The salicylic acid pathway mediates plant defense against biotrophic parasites, and because endophytes like *E. coenophiala* may be derived from biotrophic parasites, *E. coenophiala* can induce or modulate this immune response (Kari Saikkonen, Gundel, & Helander, 2013). Because the salicylic acid pathway may suppress the jasmonic acid pathway, which mediates plant defense against necrotrophic parasites (Glazebrook, 2005), *E. coenophiala* may consequently decrease resistance to necrotrophic parasites. *E. coenophiala* also produces alkaloids and other toxins (Panaccione, Beaulieu, & Cook, 2014), and these have been shown to limit growth of fungal parasites *in vitro* (Pańka, West, Guerber, & Richardson, 2013). Halliday et al. 2017 showed that inoculation

with *E. coenophiala* inhibited *C. cereale* but had no effect on *R. solani* within a host. Given that the direction and magnitude of the effect of the endophyte exerted on each parasite can vary, we hypothesized that the endophyte could alter within-host interactions between these coinfecting parasites. Specifically, because *E. coenophiala* inhibits *C. cereale,* we hypothesized that the magnitude of any facilitation or inhibition of *R. solani* by *C. cereale* would be lower when *E. coenophiala* was present.

Experimental Approach

We investigated the effects of within-host microbial interactions on the disease severity caused by a fungal parasite by conducting a series of three inoculation experiments. For each experiment, we used plants that were propagated from seed in a greenhouse. For Experiment 1 and 2, plants were grown in 993 mL pots (Deepots, D60), and for Experiment 3, plants were grown in 656 mL pots (Deepots, D40). In all experiments, plants were grown in MetroMix 360 potting mix (Sun Gro Horticulture, Agawam, MA, USA). The greenhouse temperature was kept between 19.7-22.2ºC and light was supplemented between 9am and 7pm if natural light fell below 350 W/m^2. Plants were relocated from the greenhouse to growth chambers (Percival PGC-6L; Perry, Iowa) 6 weeks after germination and given two days to acclimate. The plants were watered from the bottom of the pot during the course of the experiments. The experiments were conducted in two growth chambers with a 12-hour dark/12-hour light schedule and kept at 29°C. Humidifiers were placed on each shelf of each growth chamber and used to maintain relative humidity at 85-90%. The locations of individual plants on each shelf (2 within each growth chamber) and within each growth chamber were randomized across all treatments.

Each experiment involved varying combinations of inoculations with endophyte, *R. solani* and/or *C. cereale*. We used endophyte-inoculated and endophyte-free seed of cultivar KY-31, provided by Dr. Tim Phillips at the University of Kentucky.

Inoculations with *R. solani* involved placing a 0.6 cm PDA plug covered with actively-growing mycelia at the base of the oldest leaf of a random tiller of a plant. Successful infection of *R. solani* requires that the inoculation site be humid. To keep the inoculation site humid, we covered the PDA plug on the leaf with a piece of cotton wet with sterile water, and wrapped the base of the leaf with tin foil, which was then wrapped in parafilm. The plug, cotton, foil and parafilm were removed after two days. The *R. solani* strains used in these experiments was isolated from tall fescue in Duke Forest Research and Teaching Laboratory in summer 2015.

Inoculations with *C. cereale* occurred in one of two ways. In Experiment 2, we inoculated with mycelia in the same way as described above for *R. solani*. In Experiment 3, we inoculated with *C. cereale* spores (as described in Beirn, Wang, Clarke, & Crouch, 2015). *C. cereale* was grown in culture on PDA plates under constant light for 10 days until the *C. cereale* had filled the plates. Under sterile conditions, plates were each flooded with 10 milliliters of sterile water and a sterile loop was used to scrape the spores from the original culture. Then, the sterile loop was used to smear a new plate with spore solution. These plates were grown under constant light without being sealed for 4 days. After four days, these plates were flooded again and the resulting solution was saved. Spore concentration was adjusted to 10^6 spores/mL and 10% potato dextrose broth (PDB). To inoculate plants, 10 milliliters of spore solution were sprayed onto the plant using an atomizer, and to keep the conditions humid, the plants were covered with a plastic bag for two days. The *C. cereale* strain used in these experiments was

isolated from tall fescue in the Duke Forest Research and Teaching Laboratory in the fall of 2015 (Halliday, Heckman, Wilfahrt, & Mitchell, 2019).

Within each experiment, we also included mock-inoculated controls. When the experiment involved inoculations with PDA plugs, mock-inoculated plants were treated in the same way with sterile PDA plugs. When the experiment involved inoculation with spore suspension of 10% PDB, mock-inoculated plants were sprayed with sterile 10% PDB.

Following all experiments, plants were harvested. For experiments in which endophyte was involved, one inch-long cross-section from each of two tillers per plant were collected and frozen. The cross-sections were used to confirm the presence of the endophyte in aboveground tissue with an immunoblot assay (Agrinostics Ltd. Co, Watkinsville, GA, USA). Aboveground biomass was harvested, dried, and weighed.

Inoculation experiments

Experiment 1

To test if and how the systemic endophyte, *E. coenophiala,* affects the within-host growth of *R. solani*, we conducted an inoculation experiment in which we factorially manipulated endophyte inoculation (two levels: inoculated and mock-inoculated) and *R. solani* strain (three strains and mock inoculation). The design enabled us to investigate the potential effect of the endophyte on *R. solani,* as well as if there is any parasite intraspecific variation in this effect. Plants within each endophyte treatment (endophyte-inoculated and endophyte-free) were fully randomized within 4 parasite treatments: inoculation with one of three strains of *R. solani* or mock inoculation. Specifically, we inoculated with *R. solani* 39 endophyte-free plants (10 plants per *R. solani* strain) and 63 endophyte-inoculated plants (19 plants per *R. solani* strain). We mock-inoculated control plants (4 endophyte-free and 4 endophyte-inoculated) that received a

mock inoculation instead of inoculation with either parasite (Table 3.1). Following inoculations, the presence/absence of disease symptoms, size of lesions, and survival of leaves were observed and measured.

Of the 57 plants inoculated from seed with the endophyte, 37 plants (64.9%) were successfully infected with the endophyte in aboveground tissue, as detected with an immunoblot assay. The 21 plants in which the endophyte was inoculated but not detected in aboveground tissue were excluded from analyses. Eight mock-inoculated control plants (4 endophyte-free and 4 endophyte-inoculated) did not exhibit symptoms of either parasite, and so were excluded from analyses. These exclusions left 30 endophyte-free plants and 37 endophyte-infected plants in our analyses (Table 3.1).

Table 3.1: Experiment 1 Setup

Experimental Treatments		**Sample Size at start of experiment**	**Sample size included in analyses (after confirming endophyte presence/lack of symptoms on controls)**
Endophyte Treatment	***R. solani* treatment**		
Inoculated	Strain 1	19	10
	Strain 2	19	13
	Strain 3	19	14
	Mock	4	0
Free	Strain 1	10	10
	Strain 2	10	10
	Strain 3	10	10
	Mock	4	0

Experiment 2

To test how coinfection, as well as infection sequence, with a parasite, *C. cereale*, affects the severity of disease caused by *R. solani,* we conducted an inoculation experiment with the two parasites. Specifically, we inoculated endophyte-free plants with *C. cereale* alone, *R. solani* alone, or with both parasites. We had three treatments in which both parasites were inoculated: Simultaneous co-inoculation, sequential inoculation in which *C. cereale* was inoculated first, and

sequential inoculation in which *R. solani* was inoculated first. Inoculation of *C. cereale* was performed with mycelia on PDA plugs as described above. Specifically, 10 plants each were randomly assigned into one of six inoculation treatments (60 plants total): 1. *C. cereale* alone, 2. *R. solani* alone, 3. Simultaneous co-inoculation of *C. cereale* and *R. solani,* 4. Sequential co-inoculation: *C. cereale* first, then *R. solani,* 5. Sequential co-inoculation: *R. solani* first, then *C. cereale,* 6. Mock-inoculated control (Table 3.2).

For the sequential inoculations, the second inoculation occurred when symptoms of the first parasite appeared on all leaves in the relevant treatment. It took 10 days for *Colletotrichum* symptoms to be observed on all inoculated leaves and 2 days for *Rhizoctonia* symptoms to be observed on all inoculated. Two leaves per plant were inoculated according to the assigned treatment. Following inoculations, symptoms of *R.* solani were measured as lesion length for 2 weeks.

Ten mock-inoculated control plants did not exhibit symptoms of either *C. cereale* or *R. solani,* so were excluded from analyses. The sequential inoculation treatment in which *R. solani* was inoculated second was surveyed for damage with *R. solani* for the remaining time of the experiment, 12 days. Therefore, for analyses in which days after inoculation with *R. solani* was a predictor, we considered the first 12 days after *R. solani* inoculation for each parasite inoculation treatment (Table 3.2).

Table 3.2: Experiment 2 Setup

Experimental Treatment	Sample Size	Sample Size Included in Analyses (unless otherwise noted)
C. cereale alone	10	10
R. solani alone	10	10
Simultaneous co-inoculation	10	10
Sequential co-inoculation: *R. solani* first	10	10
Sequential co-inoculation: *C. cereale* first	10	10
Mock-inoculated Control	10	0

Experiment 3

To test if and how a systemic endophyte will shift competitive outcomes of two coinfecting parasites, we conducted an inoculation experiment in which we factorially manipulated inoculation with *E. coenophiala*, *C. cereale,* and *R. solani*. In the coinfection treatment, both parasites were inoculated simultaneously because as further discussed below, in Experiment 2, *C. cereale* had the most impact on *R. solani* within-host growth when the parasites were inoculated simultaneously. Inoculum of *C. cereale* was in the form of a spore suspension, as discussed above. This experiment involved 122 plants that were divided among 8 treatments that represented a factorial combination between endophyte inoculation (2 levels: endophyte-inoculated and endophyte-free) and parasite inoculation treatment (*C. cereale* alone, *R. solani* alone, co-inoculation of both parasites, mock-inoculated control) (Table 3.3).

Following inoculations, plants were surveyed three times per week for disease severity. Specifically, the length of each leaf and each *R. solani* lesion were measured. We also estimated the percent of leaf area infected with each parasite ("infection severity") by visually comparing leaves to reference images of leaves of known infection severity (Mitchell et al. 2002, 2003; Halliday et al. 2017). After 4 weeks, plants were harvested.

Of the 67 plants in the experimental treatments that were inoculated from seed with the endophyte, 44 (65.6%) were confirmed to have endophyte in aboveground tissue. We excluded from analyses the 23 plants that had been grown from endophyte-inoculated seed but in which endophyte was undetected. Eight control plants (4 endophyte-free and 4 endophyte-inoculated) that received a mock inoculation instead of inoculation with either parasite did not exhibit symptoms of either parasite, and so were excluded from analyses. These exclusions left 47 endophyte-free plants and 44 endophyte-infected plants (split evenly among parasite inoculation treatments) in our analyses (Table 3.3).

Table 3.3: Experiment 3 Setup

Experimental Treatment		**Sample Size at start of experiment**	**Sample size included in analyses (after confirming endophyte presence/lack of symptoms on controls)**
Endophyte Treatment	**Parasite Treatment**		
Inoculated	*C. cereale* alone	23	14
	R. solani alone	22	14
	Coinfection	22	14
	Control	4	0
Free	*C. cereale* alone	15	15
	R. solani alone	16	16
	Coinfection	16	16
	Control	4	0

Data Analysis

To summarize disease severity over time, area under the disease progress stairs (AUDPS) was calculated for each plant based on longitudinal assessments of *R. solani* disease severity using the audps function within the agricolae package (version 1.3, de Mendiburu & de Mendiburu, 2019). AUDPS estimated the integration of disease progress experienced by each leaf by adding together polygon steps between each time point (Simko & Piepho, 2011). We investigated if and how endophyte treatment (Experiment 1 and 3), parasite inoculation treatment (Experiment 2 and 3), and their interaction (Experiment 3) affected this estimate of disease

severity over time with linear mixed effects models that included an interaction term for growth chamber and shelf as a fixed effect to account for any differences due to location of plants and random slopes for leaf ID and time after *R. solani* inoculation to account for longitudinal surveying. In Experiment 2, two leaves per plant were inoculated and surveyed, so in models analyzing this data, we included leaf ID nested within plant ID as a random effect. When necessary, AUDPS was log transformed to achieve homoscedasticity and normality of residuals.

To test how parasite inoculation treatment and endophyte treatment affected *R. solani* growth rate, we used linear mixed effects models. Models included time after inoculation, parasite treatment (or strain, in the case of Experiment 1), endophyte treatment (when relevant) and all interactions as fixed effects. Additionally, we included an interaction between chamber and shelf as a fixed effect to account for any variation due to location within the experiment. We included random slopes for leaf ID and time after *R. solani* inoculation to account for longitudinal surveying. In Experiment 2, two leaves per plant were inoculated and surveyed, so in models analyzing this data, we included leaf ID nested within plant ID as a random effect. We accounted for temporal autocorrelation by including a continuous autoregressive structure of order 1 (CAR 1) in each model (Zuur et al. 2009). When necessary, lesion length was log transformed to achieve homoscedasticity and normality of residuals.

To test how parasite inoculation treatment, endophyte treatment and their interaction affect leaf survival, we performed two survival analyses. We used Cox proportional hazards models to quantify the instantaneous risk that a leaf will die at a time point given that the leaf survived to that time point. To consider how parasite inoculation treatments (4 levels: Mock-inoculated control, *C. cereale* alone, *R. solani* alone, and coinfection), we conducted a survival analyses with parasite inoculation treatment as a predictor variable. To consider the interaction

between parasite treatment and endophyte treatment, we conducted a survival analysis with just treatments involving *R. solani*. We then evaluated how the endophyte treatment, parasite treatment (single- or co-infection), and their interaction affected leaf survival. In each model, we accounted for placement within the growth chambers (chamber and shelf) using the cluster function within the survival package (version 2.43, Therneau & Grambsch, 2000).

To test how parasite inoculation treatment and endophyte treatment affect aboveground biomass, we used a linear model. The model included endophyte treatment, parasite inoculation treatment, and their interaction as fixed effects. Additionally, we included an interaction between chamber and shelf as a fixed effect to account for any variation due to location within the experiment.

Results

How is disease severity of *R. solani* affected by a systemic endophyte ?

Experiment 1 factorially manipulated endophyte inoculation and inoculation with three different *R. solani* strains. Across all tested *R. solani* strains, there was no effect of the endophyte on total disease, as estimated by AUDPS, caused by *R. solani* over the course of the experiment (Figure 3.1A, Table 3.4, $p = 0.861$). Longitudinal measurements of disease severity was also analyzed with a linear mixed effect model. While there was a significant effect of time after inoculation (Table 3.4, $p < 0.001$), indicating that lesion length increased over time, there was no significant main effect of endophyte presence ($p = 0.586$) and no significant interaction ($p = 0.648$) between endophyte presence and time after inoculation with *R. solani* (Figure 3.1B, Table 3.4). Overall, accounting for variation among strains of *R. solani,* there was no effect of the endophyte on *R. solani* disease intensity over time or *R. solani* growth rate on a leaf.

How is disease severity of *R. solani* affected by coinfection with parasite, *C.* cereale?

Experiment 2 manipulated the presence and infection sequence of *C. cereale* relative to *R. solani*. There was an 49.2-86.2% increase in total disease caused by *R. solani*, as estimated by AUDPS, in the coinfection treatments compared to when *R. solani* was inoculated on its own, though that increase had weak statistical support (Figure 3.2A, Table 3.5, ANOVA, $F=2.10$, $p = 0.117$). When modelling disease severity of *R. solani* as quantified by longitudinal measurements of lesion length with a linear mixed effects model, there was a significant interaction between coinfection treatment and time after inoculation with *R. solani*; growth rate of *R. solani* was higher on leaves in which *C. cereale* was also inoculated than when *R. solani* was inoculated on its own (Figure 3.2B, Table 3.5), indicative of the presence of a facilitative effect of *C. cereale* on *R. solani*. The extent of this increase in growth rate depended on infection sequence of the parasites on the leaf. Both the treatment in which *R. solani* was inoculated after *C. cereale* and the treatment in which *R. solani* was inoculated at the same time as *C. cereale* resulted in a growth rate of *R. solani* that was more than double that when *R. solani* was inoculated on its own (Both contrasts: $p = 0.01$). When *R. solani* was inoculated before *C. cereale,* the slope of *R. solani* lesion length over time was 50% higher than the slope when *R. solani* was inoculated on its own, but this increase with not statistically significant ($p = 0.34$).

How is disease severity of *R. solani* affected by the interaction between a systemic endophyte and parasite, *C. cereale*?

Experiment 3 factorially manipulated inoculation of *E. coenophiala* (the endophyte), and parasites, *C. cereale* and *R. solani*. Similar to Experiment 2, total disease caused by *R. solani,* as estimated by AUDPS, increased when co-inoculated with *C. cereale*. Specifically, total disease caused by *R. solani* increased by 14% when coinfected with *C. cereale* compared to when *R. solani* was inoculated on its own (Figure 3.3A, Table 3.6, $p = 0.003$). There was no significant

main effect of endophyte infection (Figure 3.3A, Table 3.6, p = 0.53) and no significant interaction between parasite co-inoculation treatment and endophyte infection (Figure 3.3A, Table 3.6, p = 0.47). In a linear mixed effects model of *R. solani* lesion length, there was a significant interactive effect of parasite co-inoculation treatment and time after inoculation with *R. solani,* with *R. solani* having a higher growth rate when co-inoculated with *C. cereale* than when it was inoculated on its own (Figure 3.3B, Table 3.6, p = 0.005). The slope of *R. solani* lesion length over time was 2.35 times higher for leaves co-inoculated with *C. cereale* than leaves inoculated with only *R. solani*. There was weak statistical support for the presence of the endophyte increasing the facilitative effect of *C. cereale* on *R. solani* (Table 3.6, p = 0.11).

With two survival analyses, we investigated if parasite inoculation and endophyte infection affected leaf survival. Leaves that were inoculated with both parasites had significantly lower survival over the course of the experiment than leaves in other parasite inoculation treatments (Figure 3.4A, Table 3.7, Wald statistic=12.34 on 3df, p = 0.006). Compared to mock-inoculated control leaves, leaves inoculated with only *C. cereale* had a 3.2 times higher hazard (p = 0.032), leaves inoculated with only *R. solani* had a 5.2 times higher hazard (p = 0.076), and leaves inoculated with both parasites had an 8.0 times higher hazard (p = 0.010). We used a second Cox proportional hazards model to assess the interaction between parasite treatment and endophyte infection that included the 4 level predictor of the factorial combination of parasite treatments involving *R. solani* (2 levels: Single Infection and coinfection with *C. cereale*) and endophyte infection (free or infected) (Wald statistic=8.59 on 3df, p = 0.04). While there was no difference in leaf survival between leaves inoculated with just *R. solani* and leaves co-inoculated with *R. solani* and *C. cereale* in endophyte-free plants (p = 0.581), co-inoculation significantly

decreased leaf survival in endophyte-infected leaves, with a 1.937x higher hazard (Figure 3.4B, Table 3.7, $p = 0.012$).

We also investigated if parasite inoculation and endophyte infection affected aboveground biomass. There was some statistical support for endophyte-infected plants that were inoculated with both parasites having lower biomass than endophyte-free plants inoculated with both parasites (Figure 3.5, $p = 0.070$). All other combinations of endophyte infection and parasite inoculation did not differ significantly in aboveground biomass.

Discussion

Higher-order interactions between microbial symbionts can have complex effects on both hosts and symbionts (Marchetto & Power, 2017). Our results demonstrate that interactions between a microbial mutualist and two coinfecting parasites can affect both parasite growth within a host and host response to coinfection. While coinfection with *C. cereale* always increased within-host growth of *R. solani*, the extent of this increase may depend on whether systemic endophyte, *E. coenophiala*, was present. Further, when the endophyte was present, leaf survival was lower and aboveground biomass was lower when *C. cereale* and *R. solani* were in coinfection compared to when *R. solani* was inoculated on its own.

The presence of the endophyte had no effect on within-host growth of parasite, *R. solani* in the absence of *C. cereale*. The mutualistic relationship between cool-season grasses and vertically-transmitted endophytes has been studied extensively. While endophyte infection has been shown to protect host plants by increasing resistance to herbivores and seed predators, as well as protect against abiotic stressors (Kauppinen, Saikkonen, Helander, Pirttilä, & Wäli, 2016; K. Saikkonen, Faeth, Helander, & Sullivan, 1998), evidence for defending against infectious disease is less consistent. While there are many studies in which such systemic grass endophytes

limit susceptibility to and progression of certain infectious diseases (Bonos, Wilson, Meyer, & Funk, 2005; C. Li, Gao, & Nan, 2007; Tian, Nan, Li, & Spangenberg, 2008), there are also studies that document no effect of these endophytes on the development of disease symptoms (L. L. Burpee & Bouton, 1993). There are still other studies that report higher susceptibility to parasites when grass hosts are infected with an endophyte (Wäli, Helander, Nissinen, & Saikkonen, 2006). One study, in particular, investigated how endophyte infection limited progression of disease symptoms caused by *Rhizoctonia zeae* in several tall fescue genotypes and found that there was a significant interaction between endophyte infection and host genotype on disease caused by the parasite (Pańka et al., 2013). This literature and our results suggest that the relationship between *Epichloë* endophytes and infectious disease is not easily generalizable, and the interactions between these endophytes and parasites may depend on parasite species, host genotype, and environmental conditions.

In contrast to the lack of an effect of *E. coenophiala* on *R. solani,* coinfection with *C. cereale* increased within-host growth of *R. solani* compared to growth of *R. solani* during single infection. Further, the magnitude of this facilitation depended on infection sequence. This effect of infection sequence is consistent with laboratory inoculation studies in other systems (Adame-Álvarez et al., 2014). *R. solani* is a necrotroph, and extracts resources by killing host cells. *C. cereale* is a hemibiotroph, with an initial asymptomatic phase in which it extracts resources as a biotroph, followed by a symptomatic phase in which it extracts resources as a necrotroph. Host plants respond to biotrophs and necrotrophs via the salicylic and jasmonic pathways, respectively, and crosstalk between these pathways can result in immune-mediated interactions between coinfecting parasites (Halliday, Umbanhowar, & Mitchell, 2018). Biotrophs are expected to facilitate necrotrophs, while necrotrophs are expected to inhibit other necrotrophs

(Glazebrook, 2005). The greatest increase in *R. solani* growth rate occurred when *C. cereale* and *R. solani* were inoculated simultaneously. Our results are therefore consistent with the initial biotrophic phase of *C. cereale* facilitating the growth of *R. solani*.

While *E. coenophiala* did not have a direct effect on the within-host growth of *R. solani,* our results suggest that the increase in growth rate of *R. solani* when in coinfection with *C. cereale* was higher in endophyte-infected plants than endophyte-free plants. Additionally, endophyte-infected leaves that were co-inoculated with both fungal parasites had lower leaf survival and endophyte-infected plants that were co-inoculated with both fungal parasites had lower aboveground biomass. While interactions among more than two species have been well-documented in free-living systems, these complex interactions have received less attention among symbionts within a shared host. Both Abbate et al. (2018) and Marchetto & Power (2017) suggest that the outcome of interactions between two parasite species can be modified by a third symbiont, whether that third symbiont be a third parasite in the case of Abbate et al. (2018) or a microbial mutualist in the case of Marchetto & Power (2017). Our results are consistent with such findings. Further, they add to support that higher-order interactions between microbial symbionts involving parasites can impact parasite growth and host response to disease.

There is mounting evidence that effects of microbial interactions under lab settings may not scale up to those under field conditions (Leung et al., 2018). In contrast to the results presented here, Halliday et al. (2017) documented an antagonistic interaction between *C. cereale* and *R. solani*. It may be that a limitation to directly extrapolating findings from lab experiments to field settings is the presence of higher-order interactions that are difficult to detect. Future work that considers higher-order interactions under field settings may clarify this connection.

We found that microbial mutualist can affect the interaction between two coinfecting parasites, having effects on the parasite and the host. As coinfections have gained interest among disease ecologists, studies of pairwise interactions among coinfecting microbes have increased but higher-order interactions are still underexplored. Consequently, while studying pairwise interactions has led to advances in our understanding of mechanisms underlying the consequences of such interactions, our results suggest that pairwise interactions may be context-dependent; they may depend on the presence of other species in the community. Higher-order interactions like these have complex effects on both hosts and symbionts, and these indirect interactions may be a mechanism that explains why pairwise interaction studies are not always be consistent with field observations. Future research must therefore consider the larger microbial community associated with a host when evaluating the implications of coinfections and move to interrogate more than pairwise interactions experimentally.

Table 3.4: Experiment 1: *E. coenophiala* vs. *R. solani*

	ANOVA: AUDPS of *R. solani*			
	numDF	**denDF**	**F**	**P**
Intercept	1	61	478.590	$p < 0.001$*
Endophyte	1	61	0.031	0.861

	ANOVA: Linear mixed effects model			
	numDF	**denDF**	**F**	**P**
Intercept	1	298	427.1674	$p < 0.001$*
DAI	1	298	117.5063	$p < 0.001$*
Endophyte	1	61	0.2997	0.5861
DAI:Endophyte	1	298	0.2092	0.6477

Table 3.5: Experiment 2: *C. cereale* vs. *R. solani*

	ANOVA: AUDPS of *R. solani*			
	numDF	**denDF**	**t**	**P**
Intercept	1	38	150.507	$p < 0.001$*
Inoc. Treatment	3	33	2.10	0.117
Chamber:Shelf	3	33		

	Linear mixed effects model of *R. solani* severity					
	Estimate	**numDF**	**denDF**	**SE**	**t**	**P**
Intercept	0.61	1	522	0.08	7.4	$p < 0.001$*
DAI	0.04	1	522	0.02	2.54	$p = 0.0115$
Inoc. Treatment (reference: single infection)		3	33			
Simultaneous	0.02			0.1	0.23	0.82
Sequential: Rhiz. First	0.05			0.1	0.6	0.55
Sequential: Rhiz. Second	-0.17			0.06	-1.61	0.12
DAI: Inoc. Treatment (reference: DAI:single infection)		3	522			
DAI: Simultaneous	0.05			0.02	2.57	0.01
DAI: Sequential: Rhiz. First	0.02			0.02	0.94	0.34
DAI: Sequential: Rhiz. Second	0.06			0.02	2.53	0.01
Chamber:Shelf		3	33			0.01

Table 3.6: Experiment 3: Factorial Manipulation of *E. coenophiala, C. cereale, and R. solani*

	ANOVA: AUDPS of *R. solani* (log-transformed)			
	numDF	**denDF**	**F**	**P**
Coinfection	1	50	9.851	0.003*
Endophyte	1	50	0.399	0.531
Coinfection:Endophyte	1	50	0.521	0.473
Chamber:Shelf	3	50		

	Linear mixed effects model of *R. solani* severity (square-root transformed)					
	Estimate	**numDF**	**denDF**	**SE**	**t**	**P**
Intercept	1.35	1	217	0.14	9.37	$p < 0.001$*
DAI	0.03	1	217	0.01	2.97	0.003*
Coinfection	0.24	1	50	0.18	1.33	0.18
Endophyte	-0.03	1	50	0.17	-0.17	0.86
DAI:Coinfection	0.02	1	217	0.02	0.87	0.005*
DAI:Endophyte	-0.02	1	217	0.02	-1.19	0.93
DAI:Coinfection:Endophyte	0.04	1	217	0.03	1.6	0.11
Chamber:Shelf		3	50			

Table 3.7: Survival Analysis Results

Leaf Mortality Risk (Cox mixed model)	**Coefficient**	**Hazard (exponentiated coefficient)**	**Coefficient SE**	**Pr(>|z|)**
Reference: control plants				
R. solani alone	1.64	5.1651	0.76	0.0763
C. cereale alone	1.17	3.22	0.6603	0.0317
Co-inoculation	2.07	8	0.8079	0.01

*Wald statistic=12.34 on 3df, p = 0.006

Leaf Mortality Risk (Cox mixed model)	**Coefficient**	**Hazard (exponentiated coefficient)**	**Coefficient SE**	**Pr(>|z|)**
Among endophyte-infected leaves Reference: single inoculation with R. solani				
Co-inoculation	0.6626	1.9397	0.4759	0.0118
Among endophyte-free leaves Reference: single inoculation with R. solani				
Co-inoculation	0.1955	1.2159	0.4372	0.581

*Wald statistic=8.59 on 3df, p = 0.04

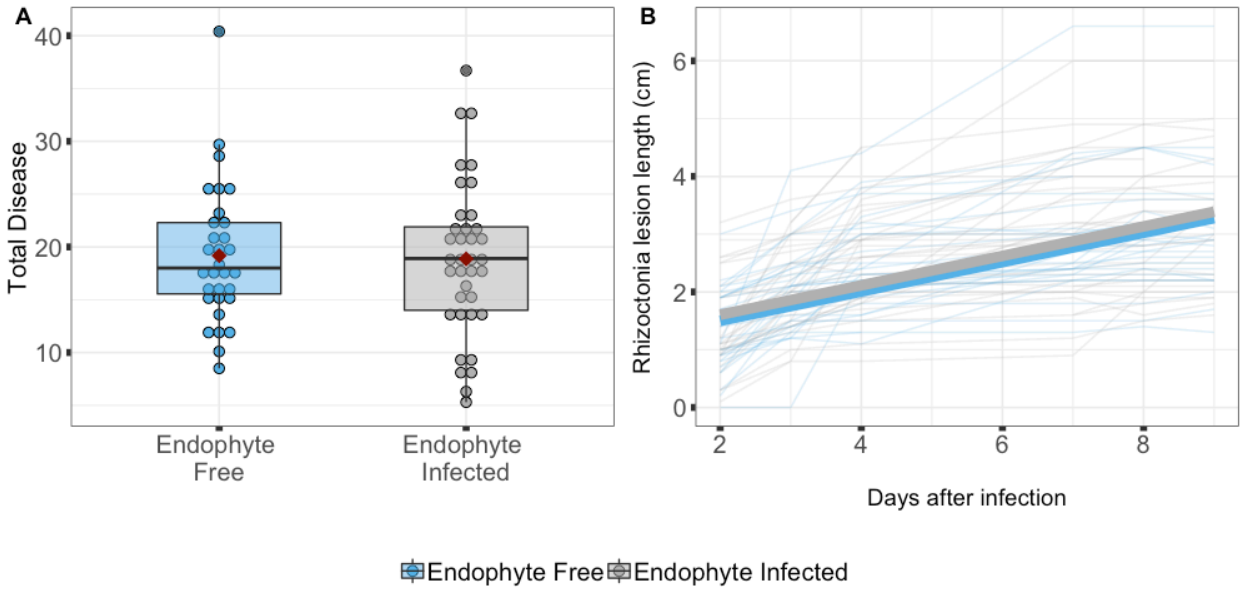


Figure 3.1: In Experiment 1, *E. coenophiala* had no effect on disease caused by *R. solani*. A. Total disease, as estimated by AUDPS with lesion length measurements over the course of the experiment, was unaffected by endophyte presence. B. Disease severity, as quantified by *R. solani* lesion length, over time. The faint are individual leaves and bolded lines are the output of a linear mixed effects model incorporating time after inoculation and endophyte presence. Lesion length significantly increased over time, but there was no significant main effect of the presence of the endophyte, as well as no significant interaction between endophyte presence and time after inoculation with *R. solani*.

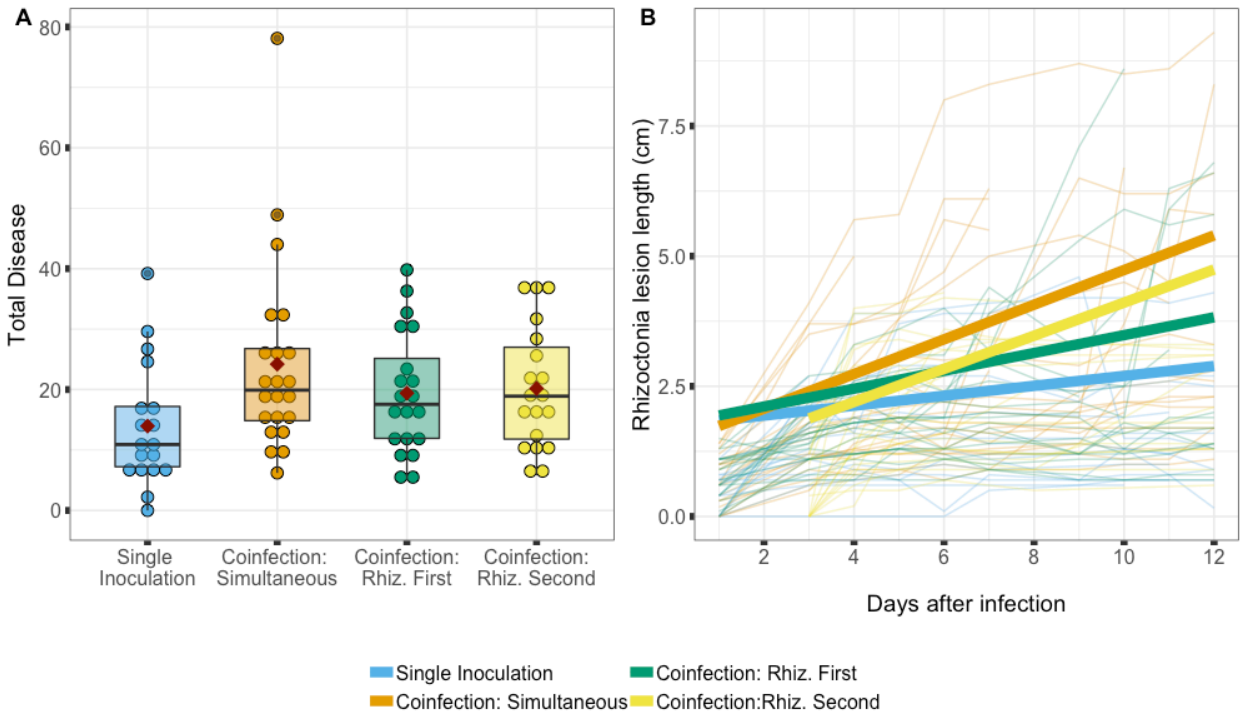


Figure 3.2: Growth rate of *R. solani* was higher when in coinfection with *C. cereale* than when inoculated on its own. A. There was a 49.2-86.2% increase in total disease caused by *R. solani*, as estimated by AUDPS, in the coinfection treatments compared to the treatment in which *R. solani* was inoculated on its own, though that increase had weak statistical support ($p = 0.117$). B. *R. solani* lesion length over time. Each faint line represents a leaf, and bolded lines are model output from a linear mixed effects model than incorporated parasite treatment and time after inoculation with *R. solani*. *R. solani* growth rate was higher in the coinfection treatments in which *C. cereale* and *R. solani* were inoculated simultaneously and when *C. cereale* was inoculated first than when inoculated on its own ($p = 0.01$).

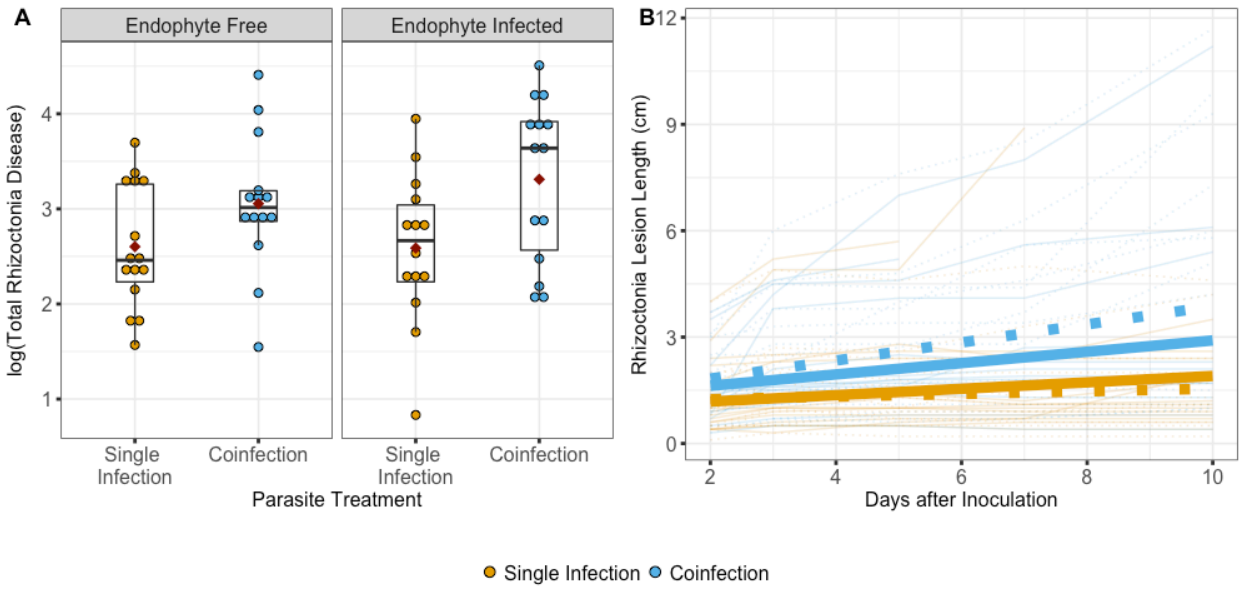


Figure 3.3: Growth rate of *R. solani* was higher when in coinfection with *C. cereale* than when inoculated on its own. A. Total disease, as estimated by AUDPS with lesion length measurements over the course of the experiment, was higher in the coinfection treatments in both endophyte-infected and endophyte-free plants. B. *R. solani* lesion length over time. Faint lines are individual leaves, and bold lines are model output from a linear mixed effects model that incorporated parasite treatment and time after inoculation with *R. solani*. Bolded dashed lines are the modeled lesion growth for endophyte-infected leaves and bolded solid lines are the modeled lesion growth for endophyte-free leaves. There was a significant interaction between time after inoculation and parasite treatment (2 levels: single infection vs. coinfection). Additionally, *R. solani* growth rate increased even more under coinfection with *C. cereale* when the endophyte is also present, but this was not a statistically significant increase.

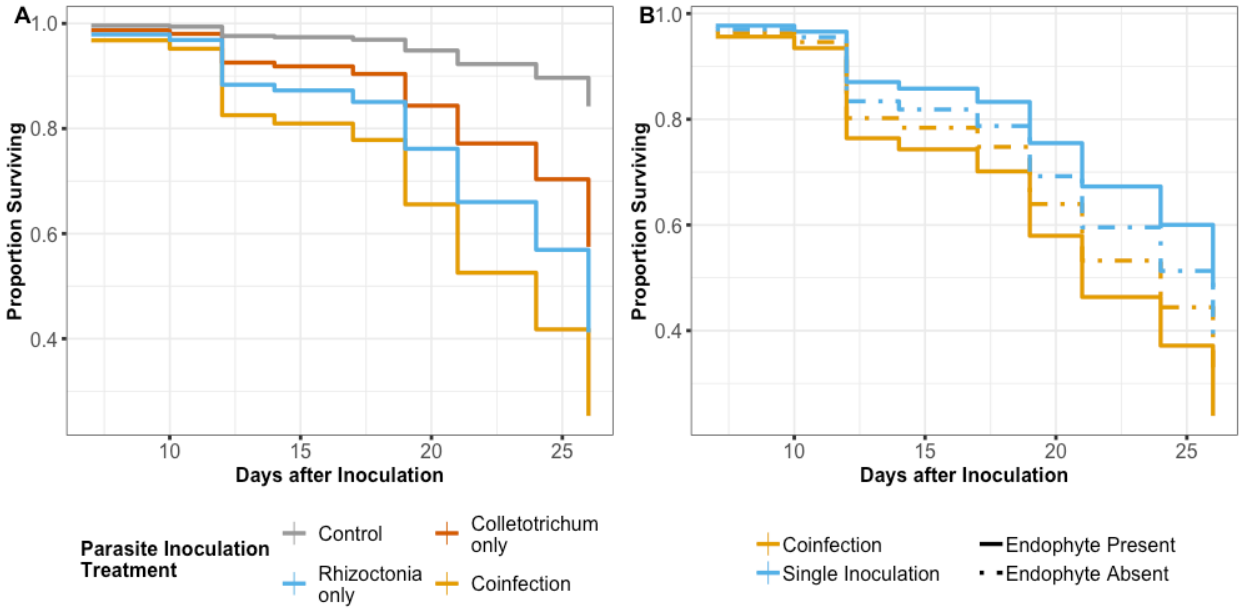


Figure 3.4: Parasite inoculation increased leaf mortality. Model-estimated relative risk of leaf death over time. Plots are results of Cox proportional hazards models. A. Parasite treatment significantly affected leaf survival. Compared to leaves that were mock-inoculated, leaves inoculated with *C. cereale* had 3.22 times higher hazard of mortality, though with weak statistical support ($p = 0.08$), leaves inoculated with *R. solani* had 5.16 times higher hazard of mortality ($p = 0.03$), and leaves that were inoculated with both parasites had an 8.00 times higher hazard of mortality ($p = 0.01$). B. While there was no difference in leaf survival between leaves inoculated with just *R. solani* and leaves co-inoculated with *R. solani* and *C. cereale* in endophyte-free plants, co-inoculation significantly increased leaf mortality in endophyte-infected leaves ($p = 0.0118$).

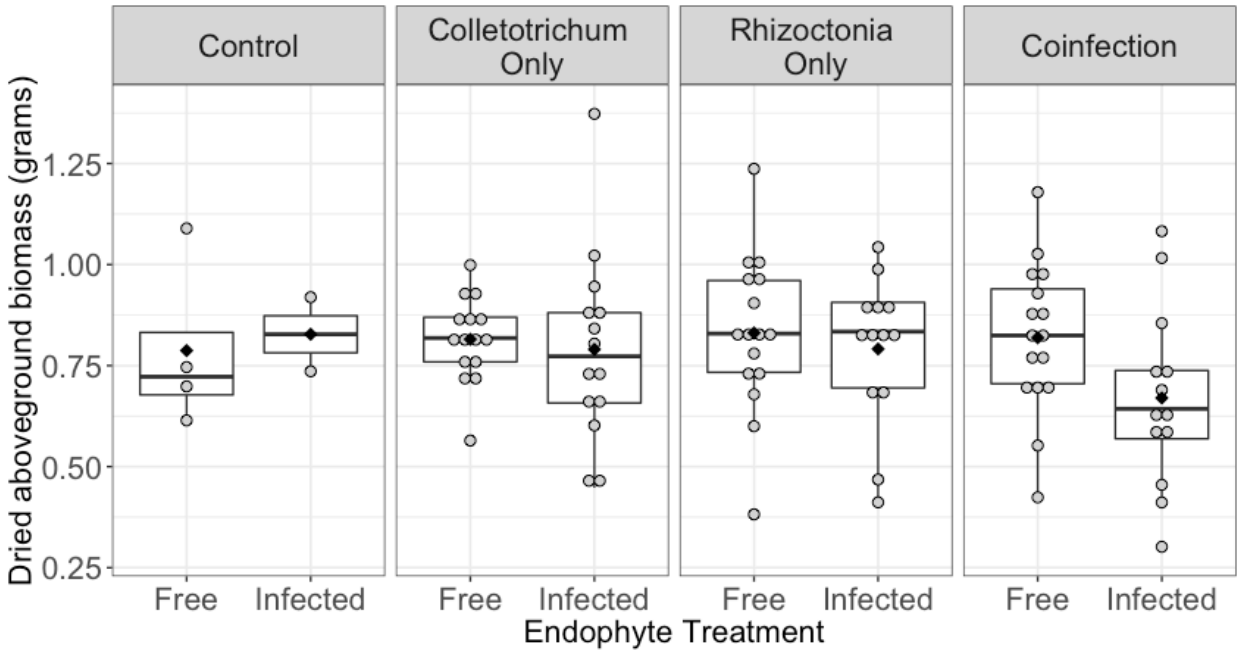


Figure 3.5: Endophyte-infected plants had lower biomass when co-inoculated with *C. cereale* and *R. solani* than all other combinations of endophyte and parasite treatments ($p = 0.07$). Points are measurements of aboveground biomass for individual plants. Boxplots show the median and the lower and upper hinges correspond to the first and third quartiles. Black diamonds show each group mean. The whiskers extend from lower and upper hinges to the smallest and largest values, respectively, no further than 1.5 times the inter-quartile range from the respective hinges, respectively. All points beyond the whiskers are outliers.

CHAPTER 4: A SYSTEMIC ENDOPHYTE AFFECTS POPULATION DYNAMICS OF INFECTIOUS DISEASE

Introduction

Parasites commonly interact with other parasites, as well as commensals and mutualists, within a shared host (Borer, Kinkel, May, & Seabloom, 2013; Tollenaere et al., 2016). In addition to the high occurrence of multiple infections of parasites, infections by defensive symbionts that protect their hosts from parasites are common (reviewed in Hopkins, Wojdak, & Belden, 2017). These defensive symbionts can also have impacts on disease at the individual- and population-levels (Hopkins et al., 2017; O'Keeffe et al., 2017). As the ubiquity of diverse within-host microbial communities have come to be realized, a challenge has been to link within- and between-host levels of related disease dynamics. Here, we address this by investigating the within- and between-host impacts of a defensive symbiont of a grass host on the severity and spread of a fungal parasite under field conditions.

Defensive symbionts can dramatically impact the survivorship, growth, and reproduction of parasites infecting the same host individual (A. E. Arnold et al., 2003; Costello, Stagaman, Dethlefsen, Bohannan, & Relman, 2012; Santhanam et al., 2015). They confer fitness benefits to host individuals by a variety of mechanisms. They may prime a host immune response to parasites (Selosse, Bessis, & Pozo, 2014) or directly interfere with an invading parasite. For example, fungal endophytes of grasses, can produce antimicrobial compounds that can reduce severity of a parasite (K. Clay, Cheplick, & Marks, 1989). Through such mechanisms, defensive

symbionts can affect host susceptibility and parasite disease severity (A. E. Arnold et al., 2003; Oliver, Moran, & Hunter, 2005).

Within-host interactions among coinfecting symbionts can have implications that scale up to populations (Cattadori, Boag, & Hudson, 2008; Telfer et al., 2010). Defensive symbionts can impact the growth and reproduction of parasites within a host and within-host accumulation is often directly or indirectly linked to between-host transmission (Tollenaere et al., 2016; Wintermantel, Cortez, Anchieta, Gulati-Sakhuja, & Hladky, 2008). Defensive symbionts may therefore be an important driver of epidemiological dynamics, which can have impacts on ecosystem function (Fisher et al., 2012). The linkage between the within-host dynamics of defensive symbionts and parasites and how that scales to impacts across host individuals remains an important frontier in disease ecology (Ezenwa & Jolles, 2015; P. T. J. Johnson, de Roode, & Fenton, 2015; Viney & Graham, 2013).

To investigate the impacts of a defensive symbiont on a parasite across levels of ecological organization, we conducted a field mesocosm experiment on a vertically-transmitted fungal endophyte, *Epichloë coenophiala,* the facultative fungal parasite *Rhizoctonia solani,* and host grass, *Lolium arundinaceum*. We performed longitudinal visual surveys of parasite damage of populations of endophyte-inoculated and endophyte-free plants following the introduction of parasite inoculum. We provide evidence that this endophyte does have a mutualistic relationship with the host at the host-individual level, but that it may not be a defensive mutualist with respect to this parasite's disease progression and spread.

Methods

Study System and Site

This experiment focused on two common fungal symbionts of the host, tall fescue (*Lolium arundinaceum*): parasite, *Rhizoctonia solani* and the vertically-transmitted systemic endophyte, *Epichloë coenophiala*. *R. solani* is a generalist parasite that causes disease on many plant species, including brown patch of tall fescue. It is a facultative parasite, as it does not rely on a plant host to survive and can persist in the soil as a saprobe. As a necrotroph, it kills living host cells and extracts resources from dead tissue. In contrast, *E. coenophiala* is considered a mutualist of tall fescue. *E. coenophiala* can increase its host's tolerance of and recovery from drought (Malinowski & Belesky, 2000) and produces bioactive alkaloids that can protect plants against herbivores (K. Saikkonen et al., 1999). In the context of infectious diseases, *E. coenophiala* can facilitate or suppress infection by fungal parasites via production of toxins and changes in host immunity, depending on parasite feeding strategy (Potter 1980, 1982; Liu et al. 2006; Saikkonen et al. 2013). Because the salicylic acid pathway may suppress the jasmonic acid pathway, which mediates plant defense against necrotrophic parasites (Glazebrook, 2005), *E. coenophiala* may consequently decrease resistance to necrotrophic parasites. *E. coenophiala* also produces alkaloids and other toxins (Panaccione et al., 2014), and these have been shown to limit growth of fungal parasites *in vitro* (Pańka et al., 2013). Empirical evidence for the direction of the interaction between *Epichloë* endophyte and *Rhizoctonia* parasites varies (Halliday, Umbanhowar, & Mitchell, 2017; Pańka, West, Guerber, & Richardson, 2013, O'Keeffe unpublished data).

This experiment was conducted at Widener Farm, an old field of the Duke Forest Teaching and Research Laboratory (Orange County, NC, USA), during the summer of 2018.

This old field produced row crops until 1996 and since 1996, has been mowed to produce hay. During the 2013-2017 growing seasons, surveys of the tall fescue population at this site showed that symptoms from this parasite began appearing on leaves in June or July, peaked in prevalence in August and September, and decreased in prevalence over the fall months (Halliday, Umbanhowar, & Mitchell, 2017).

Experimental Design and Setup

To evaluate the effects of the endophyte on the spread of the parasite through a population of tall fescue, we set up a field mesocosm experiment. Because the peak of the parasite epidemic at this site occurred in August in previous years (Halliday, Umbanhowar, et al., 2017), we conducted our experiment during that time. We contained each population within a 45-inch (1.14 meter) diameter plastic wading pool (Summer Escapes) to limit *R. solani* inoculum coming from the environment. Each population fell into one of two treatments: endophyte-inoculated or endophyte-free. We planted a total of 26 populations: 15 endophyte-inoculated and 11 endophyte-free. 2 populations in each endophyte treatment (4 total) were not inoculated with the parasite and served as controls.

Each population consisted of 13 plants: 1 plant in the center of the population (which would ultimately be inoculated with the parasite), and 12 plants surrounding the central plant at distances of 12cm, 24cm, and 36cm away (4 plants at each distance, Figure 4.1A). We used endophyte-inoculated and endophyte-free seed of cultivar KY-31, provided by Dr. Tim Phillips and Dr. Rebecca McCulley at the University of Kentucky. The 338 plants within the experiment were propagated from seed on June 25, 2018 and grown in a greenhouse for 6 weeks. The greenhouse temperature was kept between 19.7-22.2ºC and light was supplemented between 9am and 7pm if natural light fell below 350 W/m^2. All plants except for the central plants were

transplanted into the contained field mesocosm experiment on Monday, August 6, 2018, 6 weeks after germination. The plants were given four days to acclimate to the field prior to the introduction of the parasite.

All central plants across all populations were endophyte-free. Populations otherwise consisted of plants belonging to the same endophyte category (all endophyte-inoculated or endophyte-free). Plants were randomly assigned to one of the populations, and locations of the plants within the populations were also randomized. The populations were fully randomized in a 2 x 13 layout, with narrow paths separating populations (Figure 4.1B).

Plants were inoculated on August 8, 2018 with a strain of the parasite that was isolated from a tall fescue plant in 2016 in the same field as this experiment. This strain was isolated from a leaf lesion. Once in axenic culture, plugs of the leading edge of the culture were stored in mineral oil and potato dextrose broth in a -80C freezer. We plated these plugs on potato dextrose agar and the resulting growth served as the source of inoculum for this experiment. Inoculum consisted of a 6mm diameter plug of potato dextrose agar with the leading edge of the parasite culture placed directly at the base of a leaf. Parasite infection success depends on a humid environment. To maintain moisture at the site of inoculation, a cotton ball wet with sterile water was placed over the inoculum, secured with tin foil and parafilm. The plants were placed in a growth chamber (Percival PGC-6L (Perry, Iowa)) for two days with a 12 hour light/12 hour dark cycle set at 28˚C and humidity was maintained at approximately 95% with humidifiers (Vicks V5100-N Ultrasonic Humidifier) on each shelf of the growth chambers. In addition to parasite-inoculated plants, four plants were mock-inoculated with plugs of potato dextrose agar without *R. solani* mycelia. After two days, all plants inoculated with *R. solani* exhibited parasite symptoms and were transplanted into the field mesocosm experiment on August 10, 2018.

Control populations, two endophyte-inoculated and two endophyte-free, served to ensure that no environmental sources of *R. solani* contaminated the experiment, and one of each mock-inoculated plants were transplanted into these populations.

Data Collection

Following the placement of the central, parasite-inoculated plant, twice a week, for four weeks, 7 random leaves per plant were selected and observed for the presence or absence of damage caused by the parasite, as well as any other parasite damage. Each leaf was surveyed for the presence of any damage caused by parasites, herbivory, or abiotic sources. Over the course of the experiment, eight presence/absence surveys were conducted.

To measure disease severity over time, percent leaf area damaged by the parasite was quantified on individual leaves on one tagged tiller per plant once a week. On each leaf, the initial date of symptomatic infection by the parasite was recorded, and severity was estimated by visually comparing leaves to reference images of leaves of known infection severity (Halliday, Umbanhowar, et al., 2017; Mitchell, Tilman, & Groth, 2002, 2003). Over the course of the experiment, three severity surveys were conducted.

At the conclusion of the experiment, we collected and froze (-20°C freezer) inch-long cross-sections of two tillers per plant to confirm endophyte presence. Endophyte infection was tested via immunoblot (Agrinostics Ltd. Co, Watkinsville, GA). Additionally, aboveground biomass was harvested, dried and weighed.

Data Analysis

Endophyte-inoculated seed did not always result in endophyte infection in aboveground tissue. Overall, we detected the endophyte in aboveground tissue in 42% of endophyte-inoculated plants. This resulted in variation in endophyte prevalence among the endophyte-

inoculated populations. We therefore analyzed our data in two separate ways: with endophyte treatment (2 levels: endophyte-free or endophyte-inoculated) as a predictor, or with endophyte prevalence (continuous variable from 0-1) as a predictor. Additionally, control populations did not exhibit symptoms of the parasite, confirming that containment of populations limited environmental sources of inoculum.

Leaves were analyzed as host individuals because each parasite infection is restricted to a single leaf (as done by Halliday, Umbanhowar, & Mitchell, 2017). To summarize disease progression over time, area under the disease progress stairs (AUDPS) was calculated for each population using the audps function within the agricolae package (version 1.3, De Mendiburu 2016). AUDPS estimated the integration of the development of disease progress experience by each population by adding together polygon steps between each time point (Simko & Piepho, 2011). Disease intensity at each time point was calculated separately with two types of data: data from the weekly severity surveys, represented by the average leaf area damaged across all leaves surveyed within a population and data from the biweekly prevalence surveys, represented by the parasite prevalence. We investigated if and how endophyte treatment affected these estimates of disease progression over time with a linear model. Where noted, AUDPS was log transformed to achieve homoscedasticity and normality of residuals.

To further evaluate the magnitude of epidemics, we investigated longitudinal measurements of parasite prevalence using linear mixed effects models. Data on proportion of leaves infected with the parasite were logit-transformed to achieve homoscedasticity and normality of residuals. Using the nlme package (version 3.1) for linear mixed effects models, we modeled parasite prevalence within a population at a given time with a linear mixed effects model that included endophyte inoculation treatment and a third order polynomial of days after

infection, as well as their interaction, as predictors (Pinheiro, Bates, DebRoy, Sarkar, & Team, 2013). We included random slopes to account for repeated measures of the populations.

We additionally evaluated if endophyte inoculation treatment affected peak parasite prevalence. We quantified peak parasite prevalence as the highest proportion of leaves infected with the parasite at a given time point at the population-level. We tested if the endophyte inoculation treatment affected peak parasite prevalence with a linear model that included endophyte inoculation treatment as the predictor.

To evaluate disease progression within individual hosts, we analyzed disease severity of individual leaves tracked over time. We analyzed a subset of leaves that became infected with the parasite. We tested how disease severity over time was affected by endophyte status with linear mixed effects models with endophyte status (3 levels: endophyte-free, endophyte-infected, and endophyte-inoculated but not detected), time, their interaction and distance from the central, parasite-inoculated plant as predictors and accounted for repeated measures of individual leaves with random slopes. When analyzing all leaves, time was weeks after infection; when analyzing the subset of leaves that would eventually become infected with the parasite, time was weeks after infection was first observed. We log-transformed estimated disease severity to achieve homoscedasticity and normality of residuals. The emtrends function in the emmeans package (version 1.3.2, Lenth, 2018) was used to estimate and compare the slopes of time versus disease severity for each endophyte category.

Under density-dependent transmission, the contact rate between susceptible and infected individuals depends on the population density; transmission rate therefore increases with density. To investigate if host density (here, the number of leaves on a plant) significantly correlated with biomass, we used data from a field mesocosm experiment in 2017. This experiment involved ten

endophyte-inoculated populations and ten endophyte-free populations, with two of each endophyte treatment serving as controls. We investigated this potential correlation at two levels: the individual plant-level and the population level. At the individual plant-level, we used a linear mixed effects models to investigate if there was a correlation between plant-level biomass measured at the end of the experiment and the total number of leaves surveyed on a plant at the end of the experiment. We included population as a random effect. At the population level, we used a linear model to investigate if there was a correlation between biomass of all plants within a population and the total number of leaves surveyed in a given population at the end of the experiment. We then used aboveground biomass as a proxy for host density of each population. To determine if any variation in parasite peak prevalence between endophyte inoculation treatments was due to changes in host density, we analyzed the peak proportion of leaves infected with the parasite at the population-level during the experiment with pool-level biomass and endophyte population-level treatment as explanatory variables.

Results

Endophyte-inoculated populations experienced higher disease intensity over time, both when AUDPS was estimated with parasite prevalence and when estimated with disease severity data, though these differences had weak statistical support. Specifically, endophyte-inoculated populations experienced 8.3% higher disease intensity over time when estimated with parasite prevalence data (Figure 4.2A, Table 4.1, $p = 0.098$) and 54.2% higher disease intensity over time when estimated with disease severity data (Figure 4.2B, Table 4.1, $p = 0.097$). When looking at longitudinal data of prevalence over time, the endophyte inoculation treatment had no effect (Figure 4.3A, Table 4.2). Parasite prevalence was significantly predicted by days after inoculation ($p < 0.001$), but there was no main effect of the endophyte treatment ($p = 0.15$) and

no interactive effect of endophyte treatment and days after inoculation ($p = 0.24$). When endophyte prevalence (a continuous variable), rather than endophyte treatment (a categorical variable), was used as a predictor endophyte infection similarly had no effect on parasite prevalence over time (results not shown).

While endophyte treatment did not significantly affect disease intensity over time, it did significantly affect the peak parasite prevalence, as the prevalence of endophyte-inoculated populations (mean prevalence = 0.43) had a 27% higher peak prevalence than endophyte-free populations (mean prevalence = 0.34) (Table 4.3, Figure 4.3B, $p = 0.007$).

We investigated if there was any effect of endophyte infection on disease severity over time using the weekly estimates of disease severity (percent leaf damaged) that were made of individual leaves over time. Among leaves in which damage from the parasite was observed, there was a significant effect of time ($p < 0.0001$). Additionally, there was an interaction between time and endophyte infection category with weak statistical support. Specifically, the slope of disease severity versus time among endophyte-infected leaves was approximately double that of endophyte-free leaves (Table 4.4, Figure 4.4, $p = 0.0807$).

At the individual plant level, endophyte infection significantly affected biomass (Figure 4.5A, Figure B4.1, Table 4.5). Endophyte-infected plants had the highest aboveground biomass, and there were no differences in biomass between plants grown from endophyte-absent seed and plants in which endophyte was not detected in aboveground tissue but grown from endophyte-inoculated seed. Specifically, endophyte-infected plants had 29.7% higher aboveground biomass than endophyte free plants, on average, and 19.1% higher aboveground biomass than plants in which endophyte was not detected in aboveground tissue but grown from endophyte-inoculated seed. We also investigated how these differences scaled up to the population level. The biomass

of endophyte-inoculated populations tended to be higher than that of populations in which the endophyte was not inoculated, but this difference was not statistically significant (Figure 4.5B, Table 4.5, $p = 0.14$).

To determine if biomass could be a proxy for host density, we investigated if there was a correlation between biomass measured at the end of the experiment and the total number of leaves surveyed at the end of the experiment. We found that at the plant level, biomass and the number of leaves were significantly positively correlated (Figure 4.6A, Table 4.6, Marginal $R^2=0.21$, $p < 0.001$). Given that parasite prevalence measures were at the population level, we also considered the relationship between population-level biomass and population-level host density. We aggregated the biomass and number of leaves of all the plants within each population, and while there was weaker statistical support, our data suggests that the positive correlation between biomass and host density held at the population-level as well (Figure 4.6B, Table 4.6, $R^2=0.198$, $p = 0.08$).

Given this correlation between biomass and host density, we used population-level biomass as a proxy for host density and tested if the effect of endophyte treatment on peak prevalence was driven by higher host densities within endophyte-inoculated populations. Biomass and endophyte treatment explained approximately 56% of variance in peak parasite prevalence. Biomass significantly positively correlated with peak parasite prevalence ($p = 0.006$), and independent of this effect, endophyte inoculation increased peak prevalence by 23.5%, with mean peak prevalence of endophyte-free populations at 0.34 and mean peak prevalence of endophyte-inoculated populations at 0.42 ($p = 0.01$) (Table 4.7, Figure 4.7).

Discussion

Our study provides experimental evidence that within-host disease dynamics change under coinfection with a systemic endophyte, resulting in an increase in population-level parasite peak prevalence. Specifically, we investigated the responses of parasite growth within a host individual and parasite spread through a host population to the presence of a systemic endophyte under field conditions. We found that populations of tall fescue inoculated with endophyte, *E. coenophiala*, experienced a higher peak prevalence of parasite, *R. solani*, over the course of the experimental epidemic.

The mutualistic relationship between cool-season grasses and vertically-transmitted endophytes has been studied extensively. We found that infection with *E. coenophiala* resulted in an increase in aboveground biomass. While *Epichloë* endophyte infection has been shown to protect host plants by increasing resistance to herbivores, and seed predators, as well as protect against abiotic stressors (Kauppinen et al., 2016; K. Saikkonen et al., 1998), evidence for defending against infectious disease is less consistent. In this experiment, within a host, there was no support for *E. coenophiala* limiting disease progression. There was, however, some statistical support for this endophyte facilitating the growth rate of *R. solani* within a leaf. Our results are consistent with studies in which no effect (L. L. Burpee & Bouton, 1993; Halliday, Umbanhowar, et al., 2017) or a facilitative effect of the endophyte on a parasite was identified (Halliday, Umbanhowar, et al., 2017; Wäli et al., 2006). Vertically-transmitted fungal endophytes can impact fungal parasites via resource competition and changes in host immunity, which depends on parasite feeding strategies (Kari Saikkonen et al., 2013). We expected that *R. solani,* as a necrotroph, would be inhibited by *E. coenophiala,* but the direction of the effect of these endophytes on parasites likely depends on host genotype and environmental conditions

(Krauss et al., 2007; Pańka et al., 2013), and further experimentation is needed to determine the mechanism underlying the potential facilitation of *R. solani* by *E. coenophiala*.

Within host populations, there was no effect of *Epichloë* inoculation on parasite spread over time, but we did find that endophyte inoculation increased peak parasite prevalence experienced by a population. We hypothesized that under density-dependent transmission, the benefit to the host individual of increased aboveground biomass, which in this case, correlates with host density (here, number of leaves), may have resulted in a higher contact rate between hosts and consequently, higher parasite peak prevalence. While biomass and peak parasite prevalence were significantly positively correlated, consistent with density-dependent transmission, our results suggest that *Epichloë* impacted peak parasite prevalence beyond effects of biomass.

There is growing evidence that the direction and magnitude of the consequences of within-host microbial interactions are strongly affected by environmental factors (Halliday, Umbanhowar, et al., 2017; Leung et al., 2018). Our study provides an important contribution to this understanding, as it expands upon previous work investigating the interaction between this hypothesized mutualist and this parasite under controlled settings (O'Keeffe, Simha, Mitchell, unpublished) by interrogating the impacts of the same interaction on parasite growth within a host individual and parasite spread at the population level under field conditions. Foliar fungal parasites have been studied extensively and can serve as a suitable model system to investigate microbiome/parasite interactions under field settings. Here, we show that field mesocosm experiments offer the ability to investigate the effect of within-host microbial interactions on parasite growth and spread.

Within-host microbial interactions can influence natural epidemics in complex ways (Halliday, Umbanhowar, et al., 2017; Mordecai, Gross, & Mitchell, 2015). Within hosts, we found evidence that a mutualist that is often considered a defensive symbiont actually likely facilitated the within-host growth of this parasite. Across hosts, we found that populations inoculated with this mutualist experienced higher peak prevalence of this parasite. These results demonstrate that within-host interactions among parasites and non-pathogenic microbes can impact epidemic dynamics, and we propose that field mesocosm experiments can yield important insight into disease dynamics across populations under field settings.

Table 4.1: Endophyte-inoculated populations tended to have heavier disease over course of experiment, on average, though this difference had weak statistical support.

	AUDPS: Parasite Prevalence					AUDPS: Disease Severity			
	F	numDF	denDF	P		F	numDF	denDF	P
Endophyte Inoculation	3.02	1	20	0.098		3.05	1	20	0.097

Table 4.2: There was no significant effect of endophyte inoculation on parasite prevalence over time.

	numDF	**denDF**	**F**	**p**
Fixed Effects				
Days After Inoculation (third order polynomial)	3	148	30.87	<0.0001
Endophyte Status	1	20	2.2115	0.1534
Days After Inoculation:Endophyte Status	3	148	1.4091	0.2427

Table 4.3: There was a significant effect of endophyte on peak prevalence

	numDF	**denDF**	**F**	**p**
Endophyte Status	1	20	9.218	0.006795

Table 4.4: Disease severity over time as determined by endophyte infection

	numDF	**denDF**	**F**	**p**
Weeks After Infection	1	176	41.1052	<0.0001
Endophyte Status	2	83	1.2167	0.3014
Distance from Inoculated Plant	1	83	1.3478	0.249
Weeks After Infection:Endophyte Status	2	176	2.554	0.0807

Table 4.5:Endophyte infection increased host aboveground biomass.

Plant-Level Biomass	**numDF**	**denDF**	**F**	**p**
Endophyte Status	2	262	4.1283	0.0171
Population-Level Biomass	**numDF**	**denDF**	**F**	**p**
Endophyte Status	1	20	2.3546	0.14

Table 4.6: Aboveground biomass correlated with number of leaves at the individual plant- and population-level

	Individual Plant Level				Population Level			
	F	numDF	denDF	P	F	numDF	denDF	P
Biomass	63.7751	1	174	<.0001	3.4715	1	14	0.084

Table 4.7: There was a significant effect of endophyte on modelled peak prevalence, even after accounting for effect of biomass on prevalence

	numDF	**denDF**	**F**	**p**
Endophyte Status	1	20	12.6881	0.01
Biomass	1	20	9.126	0.006
Endophyte:Biomass	1	20	0.02	0.87

Figure 4.1: Field Mesocosm Experimental Layout. A. Plant locations within each population. Each population was comprised of 13 plants, one plant in the center that was inoculated with the parasite with 12 plants surrounding it. The surrounding plants were at three different distances away from the central plant (12 cm, 24 cm, and 36 cm). Distance away from the central plant is shown in gray scale. Distances were such that neighboring plants had leaves in contact for most of the experiment. B. Arrangement of 26 populations.

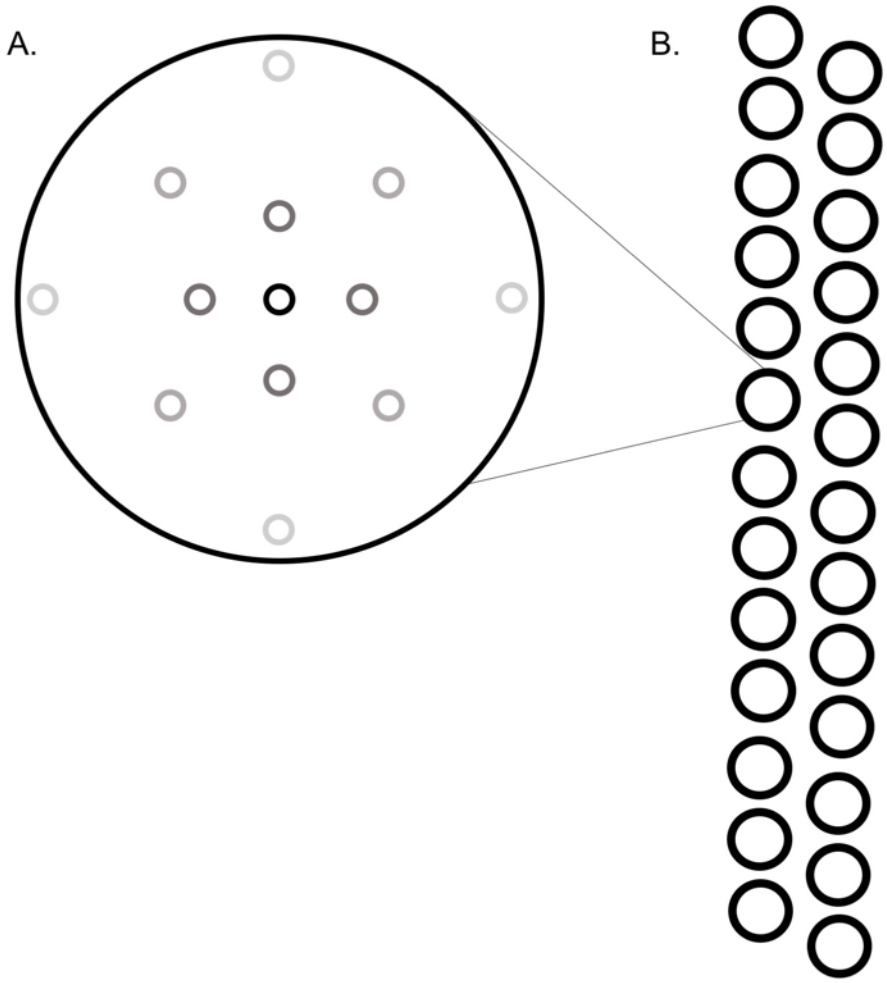


Figure 4.2 Endophyte-inoculated population tended to have heavier disease over course of experiment, on average though this difference was not statistically significant. Disease intensity over time of each population as calculated by AUDPS (log-transformed). AUDPS was calculated using two different disease measures. A. Parasite prevalence over time (8 surveys) B. Average disease severity over time (3 surveys). Boxplots show median, 25th, and 75th percentile, with dots showing data for each population replicate. Whiskers extend to the lowest and highest values no further than ± 1.5 times the inter-quartile range.

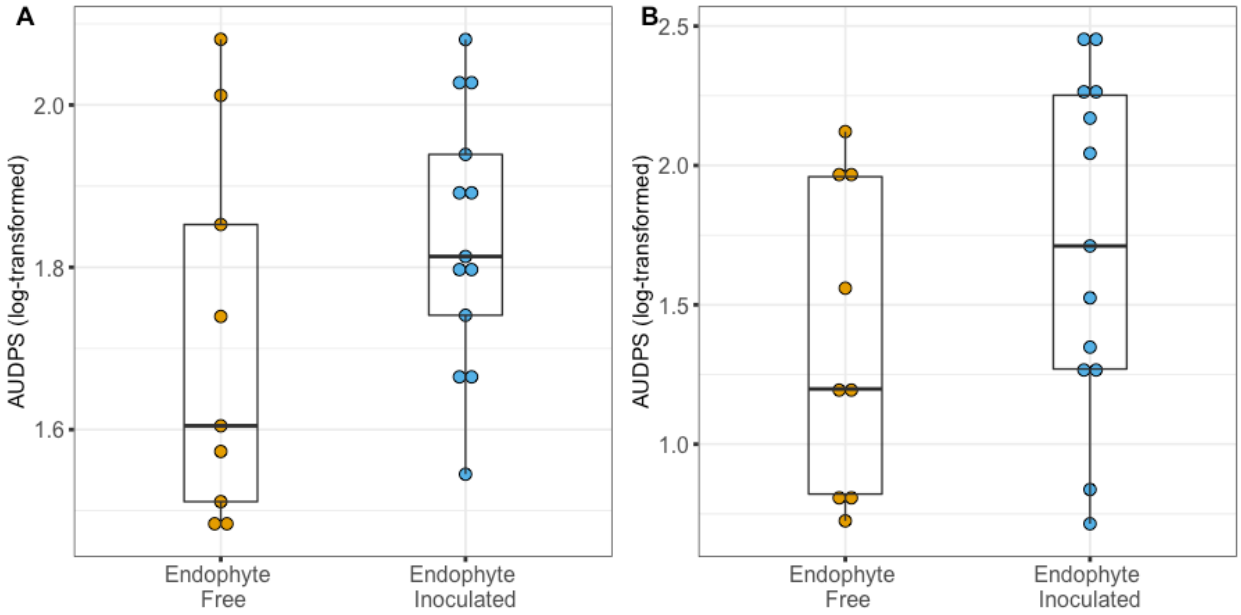


Figure 4.3 Endophyte inoculations did not affect parasite spread over time but did affect peak parasite prevalence. A. Bold lines are model-predicted means of parasite prevalence, measured as proportion of infected leaves, over the course of 4 weeks post-inoculation with the parasite for populations in which endophyte was inoculated (blue) and populations in which endophyte was absent (orange). B. Endophyte-inoculated populations had 8.6% higher peak parasite prevalence peak prevalence than endophyte-free populations. Bars indicate mean peak parasite prevalence and error bars are ± 1 SE.

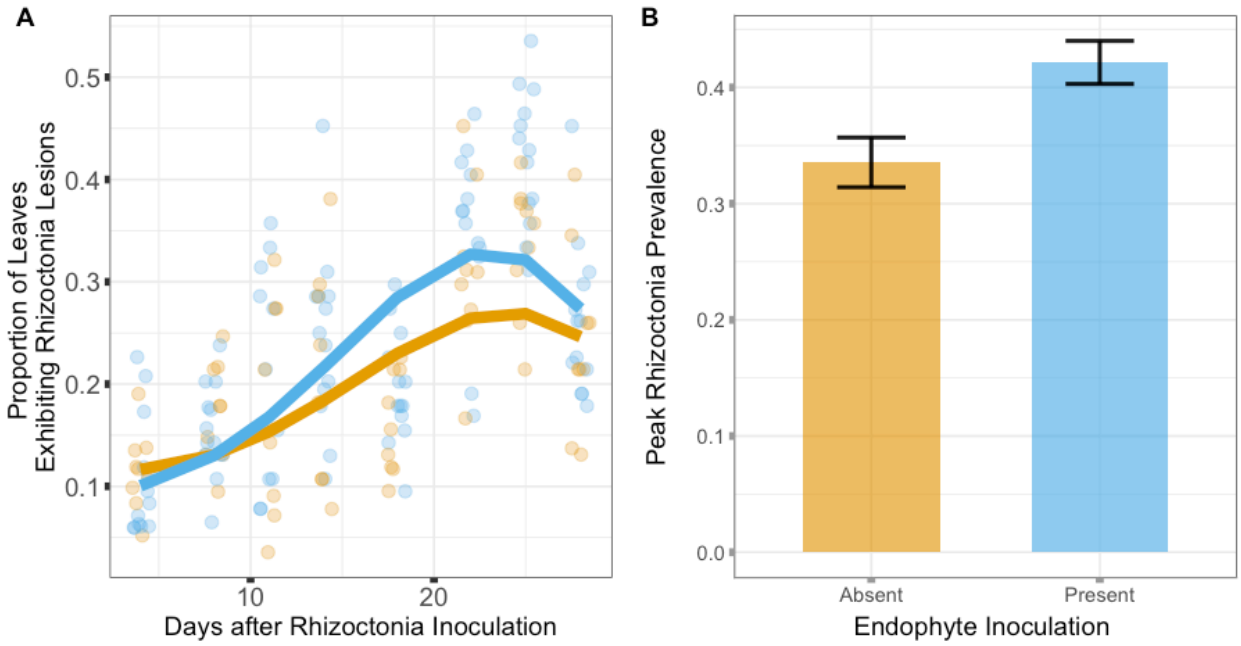


Figure 4.4: Longitudinal analysis of disease severity. Among leaves in which damage from the parasite was eventually observed, there was a significant effect of time ($p < 0.0001$) and the slopes of time and disease severity for each endophyte category differed, but that difference was trending towards statistical significance ($p = 0.0807$). Specifically, the slope of time versus disease severity among endophyte-infected leaves was approximately double that of endophyte-free leaves. Lines are model predictions and points are estimates of percent leaf damaged of an individual leaf at a given time.

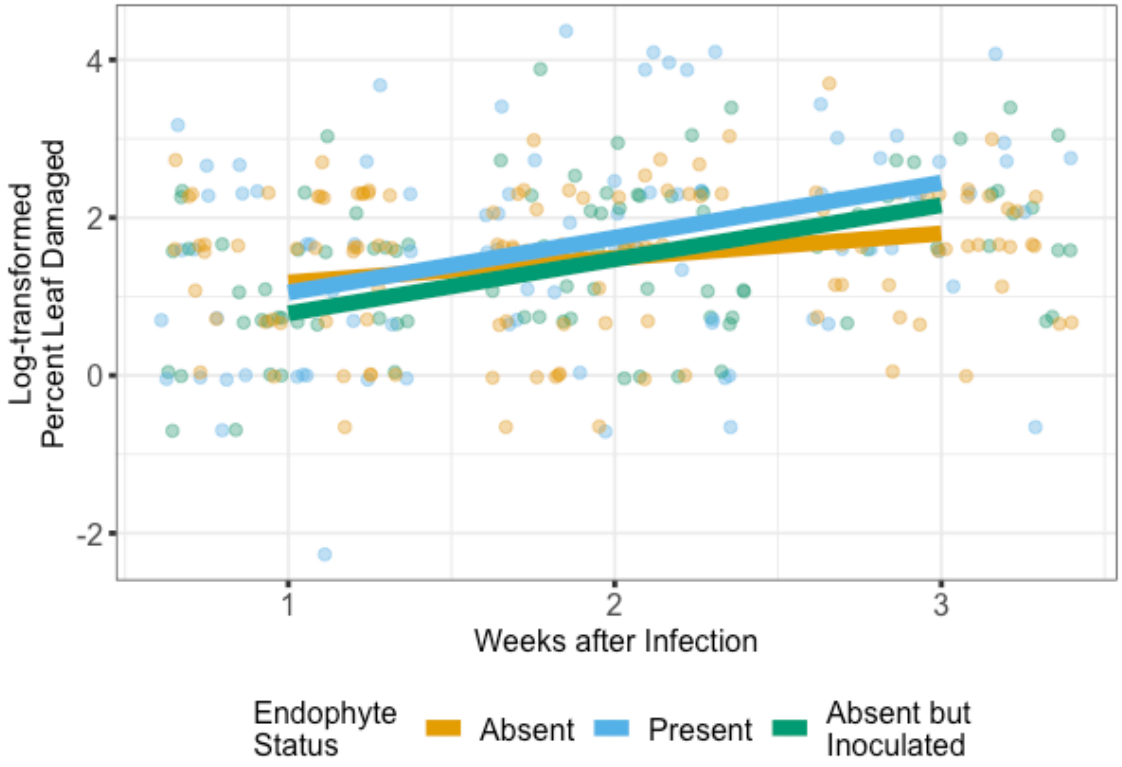


Figure 4.5 Endophyte infection increased host aboveground biomass. A. Host individuals confirmed to be endophyte-infected produced more aboveground biomass (mean = 9.83 g) than endophyte-free plants (mean = 7.37 g) or endophyte-free plants that had been inoculated with endophyte (mean = 8.36 g). Pairwise comparisons indicate statistical significance of differences between groups according to Tukey's HSD test. B. At the population level, endophyte-inoculated populations had higher total aboveground biomass, though this difference was not statistically significant. Boxplots show median, 25th, and 75th percentile, with dots showing data for each population replicate. Whiskers extend to the lowest and highest values no further than ± 1.5 times the inter-quartile range.

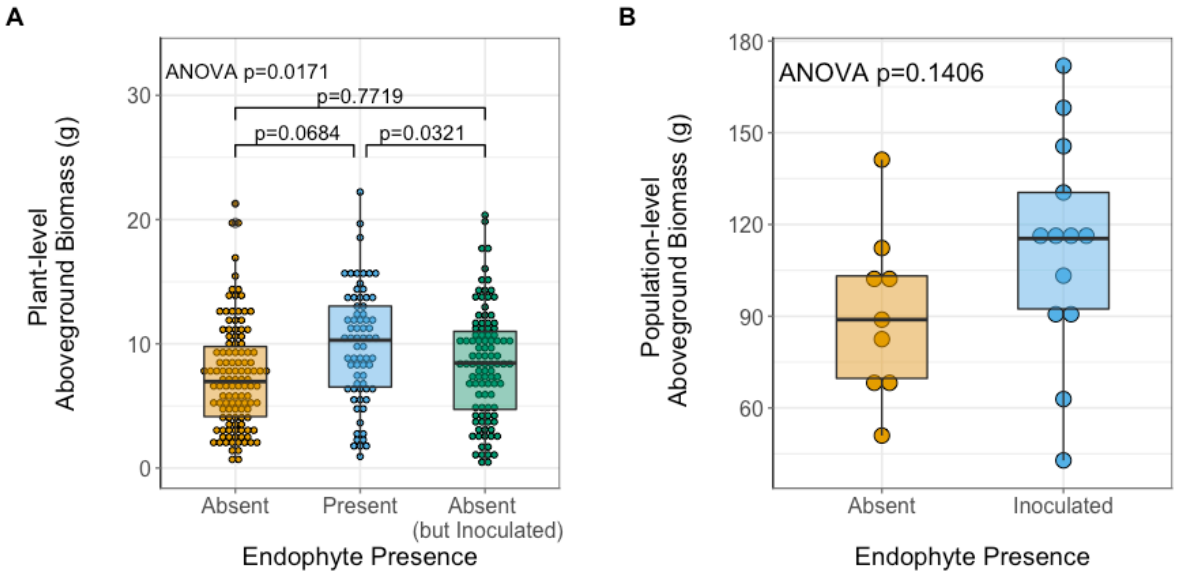


Figure 4.6 Number of leaves on a plant and in a population were correlated with dry aboveground biomass of a plant and population, respectively. A. At the plant level, we used a linear mixed effects model, with random intercepts for population to determine that the total number of leaves at the end of the experiment was correlated with dry aboveground biomass (Marginal R^2=0.21, Conditional R^2=0.42, $p < 0.001$). B. We summed the total number of leaves and the biomass of each population (the level at which other analyses were performed). The total number of leaves at the end of the experiment was correlated with dry aboveground biomass (R^2=0.198, $p = 0.08$)

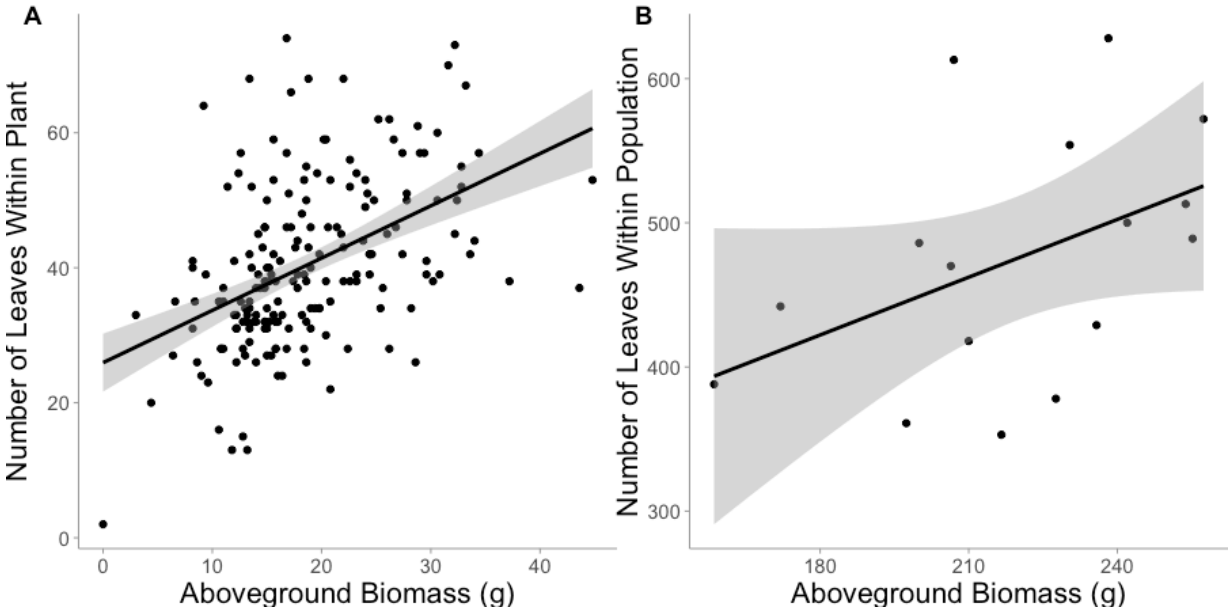
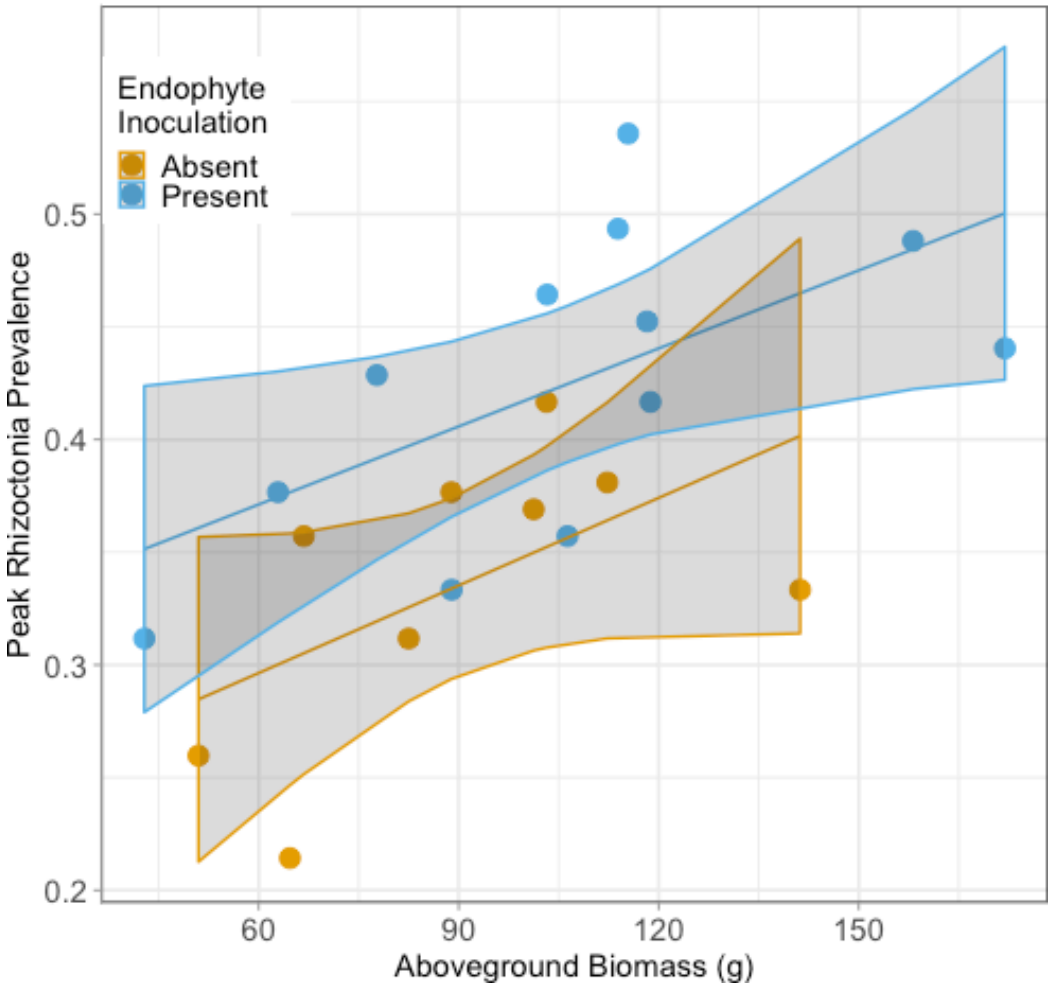


Figure 4.7 Peak Rhizoctonia Prevalence was significantly correlated with dry aboveground biomass. Each population was either inoculated with endophyte (blue) or free of endophyte (orange). Biomass significantly predicted peak parasite prevalence ($p = 0.006$), and independent of this effect, endophyte inoculation increased peak prevalence by 0.07 ($p = 0.01$).



CHAPTER 5: THE FUNGAL LEAF MICROBIOME OF A GRASS HOST UNDER NATURAL INFECTIONS BY DIVERSE PARASITES

Introduction

Parasites can affect, and be affected by, the microbiome of a host. While the microbiome of vertebrates is dominated by bacteria, fungi can play an important role in the microbiome within leaves of a plant (Christian, Whitaker, & Clay, 2015). Interactions between parasites and the components of the microbiome can have consequences for host susceptibility, disease severity, and transmission (Berg & Koskella, 2018). Investigations of the associations between parasites and the microbiome to elucidate how the microbiome influences susceptibility to infectious diseases (Libertucci & Young, 2019) and how parasites affect the microbiome (Aivelo & Norberg, 2018) are becoming increasingly common. How associations between parasites and the microbiome are modulated by variation among parasite species is still poorly understood. Here, we address this by characterizing the fungal microbiome as it associates with symptoms of three co-occurring fungal parasites that each have distinct feeding strategies.

Parasite species represent many evolutionary groups (Commission, 2006). Even within taxonomic groups, parasites vary in traits such as growth rate, generation time, and feeding strategy, and such variation can lead to differential effects on host fitness (Leggett, Cornwallis, Buckling, & West, 2017). Parasite feeding strategies have been linked to differences in stimulated host responses and effects on host condition (Budischak, O'Neal, Jolles, & Ezenwa, 2018; Glazebrook, 2005). The few parasite-microbiome studies of multi-parasite systems suggest that different parasites associate differently with the diversity and composition of a host's

microbiome (Aivelo & Norberg, 2018), but our understanding of if and how parasite traits can predict these different associations with the microbiome is poorly understood.

Microbes may interact with parasites by competing for resources, releasing antimicrobial compounds, or altering a host immune response. Such interactions can lead to consequences for host health, making the host more susceptible to, tolerant of, or resistant to parasites (A. E. Arnold et al., 2003; Busby, Peay, & Newcombe, 2016; Hayes et al., 2010). At the same time, the microbiome is dynamic and the introduction of a parasite can lead to a change in microbiome diversity and composition (Barman et al., 2008; Jani & Briggs, 2014). The link between parasites and host microbial diversity varies, with some studies showing no relationship, some studies showing a negative relationship (Jani & Briggs, 2014; Leung et al., 2018; Mosca, Leclerc, & Hugot, 2016; Wu, Stanley, Rodgers, Swick, & Moore, 2014), and still others showing a positive relationship (S. C. Lee et al., 2014). Parasites may also act as niche modifiers, as defined by Fukami (2015), by impacting the host environment in such a way as to make the host more or less suitable to new colonizers. The stability of the microbiome in the face of a parasite can have implications for consequent disease severity and host health. Given that different species house unique microbiomes (Trinh, Zaneveld, Safranek, & Rabinowitz, 2018; Turner, James, & Poole, 2013), parasite-microbiome studies of single host/multi-parasite systems may increase our understanding of parasite-specific patterns of associations with a host microbiome.

There are still some limitations to our understanding of the relationship between the microbiome and infectious disease. While there are many studies that have experimentally investigated the relationship between host-associated microbiota and parasites under controlled settings (Koch & Schmid-Hempel, 2011), fewer studies have investigated the associations between host-associated microbiota and parasites under field conditions (but refer to Jani &

Briggs, 2014). Such studies are limited by suitable model systems for exploring these questions in the field. The long-lived nature of some hosts, limited ability to detect the disease observationally in live hosts in the field, difficulty of excising infected tissue from animals and ethical concerns can be limiting. However, recent work in Leung et al. 2018 suggests that the relationships can actually be reversed under lab versus field settings. Patterns seen under controlled settings may not simply scale up under field settings.

Here, we seek to determine how the diversity and composition of the fungal microbiome within a plant host are associated with symptoms caused by three different fungal parasites under field conditions. We hypothesized that parasites that create new niches within their host (for example, by creating necrotic tissue) may lead to a change in fungal taxa richness and composition while parasites that do not create new niches within their host (for example, by keeping host cells alive while extracting resources) would not be associated with a change in fungal diversity or composition. If the case, we predicted that the composition of the fungal community associated with leaves exhibiting symptoms of parasites that kill host cells and create necrotic tissue would then have higher relative abundance of saprotrophic fungi than asymptomatic leaves. To this end, we conducted a high-throughput sequencing survey of the internal transcribed spacer (ITS) region of the fungal community of individual hosts of a population of grass, tall fescue. Specifically, we examined if and how the diversity and composition of the fungal taxa within tall fescue associated with symptoms of the three co-occurring fungal parasites.

Methods

Study System

Foliar fungal parasites can serve as a suitable model system to investigate microbiome/parasite interactions under field setting. Not only are their consequent diseases of interest, they are often readily identifiable by symptom, which facilitates observational studies in the field. Plant fungal parasites are diverse and vary in reliance on a host, feeding strategy, and stimulated host immune response. Some parasites require host tissue to remain alive while they grow within the host (biotrophs) and others kill host tissue as they spread (necrotrophs). Still others have more complex infective processes, initially infecting asymptomatically and then transitioning to killing host tissue in a second phase (hemibiotrophs). In turn, each feeding strategy stimulates a different immune defense response from the plant host (Glazebrook, 2005). As such, parasites are altering their host environment, which could in turn, impact the diversity and composition of resident microbes.

To investigate how within-host fungal diversity and composition associate with disease symptoms, we focused on a grass host, tall fescue (*Lolium arundinaceum*), and three of its fungal parasites (Table 5.1). *Colletotrichum cereale* is a rain-dispersed, hemibiotrophic parasite that infects host tissue in two stages: initially, it keeps host tissue alive asymptomatically, and then it switches to killing host tissue, which is when symptoms start appearing. *Rhizoctonia solani*, is a soil-borne, necrotrophic parasite, killing host tissue as it grows within a host. It is a facultative parasite, as it does not rely on a host to survive; it can also persist as a saprobe in the soil. *Puccinia coronata* is a wind-dispersed, obligately biotrophic parasite, keeping host tissue alive when it infects. Given that *C. cereale* and *R. solani* ultimately kill host tissue as they feed, we hypothesized that these parasites would be niche modifiers, and as such, their symptoms would

be associated with changes in fungal diversity and composition compared to asymptomatic tissue. *P. coronata* keeps host tissue alive as it feeds, so we did not hypothesize a change in fungal diversity and composition based on this symptom.

Study Site

Samples were collected at the Duke Forest Teaching and Research Laboratory. Specifically, we collected samples in a grass-dominated field within the Blackwood Division of Duke Forest in Orange County, North Carolina (35° 58'N, 79° 5'W). This site was chosen because the vegetation is primarily tall fescue (Fluxnet, 2013) and the three focal fungal parasites co-occur within this tall fescue population.

Sample Collection

Samples were collected every 20 meters along 6 parallel transects running east to west, each 100 meters long and spaced approximately 20 meters apart (Figure 5.1). Samples were collected over the course of two days in late September 2016. At each site along the transects, we estimated the density of tall fescue within a 0.5m x 0.5m quadrat and quantified the number of grass and non-grass species present within that quadrat. Within each quadrat, we also randomly selected 5 tillers of tall fescue and quantified the presence/absence of symptoms of *Rhizoctonia solani* on each leaf on each tiller. When there were fewer than 5 tillers of tall fescue within the quadrat (occurring in 4 quadrats), we surveyed all tillers within the quadrat.

At each site, we collected 4 whole leaves—one leaf with symptoms of infection by *R. solani*, one leaf with symptoms of *C. cereale*, one leaf with symptoms of *P. coronata*, and one asymptomatic leaf. Because we collected one sample of each symptom at each site, samples from the same parasite were collected at least 8 meters apart. This minimum distance would minimize the effect of spatial autocorrelation (A. E. Arnold, Henk, Eells, Lutzoni, & Vilgalys, 2007;

Higgins, Arnold, Miadlikowska, Sarvate, & Lutzoni, 2007). To standardize the approximate age of sampled leaves, we only sampled the oldest fully-expanded non-senescing leaf on a tiller. Leaves were haphazardly collected, and for each leaf exhibiting disease symptoms, we estimated the percent of leaf area infected with each parasite (infection severity) by visually comparing leaves to reference images of leaves of known infection severity (e.g., Mitchell et al. 2002, 2003; Halliday et al. 2017).

From each leaf, we cut approximately uniform segments. For the leaves that exhibited disease symptoms, we cut two separate segments spaced at least 10 centimeters apart: 1 asymptomatic segment and 1 symptomatic segment. We stored each sample in individual plastic bags that were then placed on ice and processed back in the laboratory within 4 hours. For the asymptomatic leaves, we cut the same size segment from the leaf and stored and processed in the same way.

Sample Preparation, DNA extractions, and Sequencing

All samples were brought back to the lab and processed on the same day as collection. Each sample was photographed with a ruler and weighed. Leaf segments were washed under running DI water for 30 seconds. Following surface treatment, samples were stored in -80C freezer.

We assayed the fungal communities of tall fescue leaves by sequencing the internal transcribed spacer (ITS) region. The ITS region is a region of the nuclear ribosomal RNA cistron that is often used as a DNA barcode for fungi, as it has a clearly defined barcode gap between inter- and intraspecific variation (Schoch et al., 2012). Leaf segments were ground under liquid nitrogen with a mortar and pestle and transferred to 96-well plates for DNA extraction. DNA extraction was performed with the DNEasy PowerSoil kit according to the manufacturer's

protocol (Qiagen). We amplified the first part of the internal transcribed spacer (ITS1) with a modified version of the primer set ITS1F and ITS2 for parallel sequencing on the Illumina MiSeq platform (Illumina, San Diego, CA, USA) (Smith & Peay, 2014; White, 1990). Each 25 uL PCR reaction had 2.5 uL of 10x PCR Buffer, 3.5uL of MgCl2, 1uL of ITS1-F, 1uL of ITS2, 0.5uL of dNTPs, 0.63uL of Taq polymerase, 13.12uL of water, and 3uL of DNA. The reactions were performed with the following cycle conditions: initial denaturation 95C for 1 minute followed by 40 cycles of 94C for 30 seconds, 52C for 30 seconds, 68C for 30 seconds and a final elongation at 68C for 7 minutes. We visualized PCR products using gel electrophoresis, cleaned samples with AMPure beads according to (Lundberg, Yourstone, Mieczkowski, Jones, & Dangl, 2013), and concentration-normalized (using Qubit Fluorometric Quantitation, Life Technologies, Germany). The cleaned amplicons were pooled in one run on an Illumina MiSeq instrument at the UNC High Throughput Sequencing Facility using a paired-end 2 x 250 bp kit. A spike of 10% PhiX was added to the library to increase sample heterogeneity. All raw sequence data will be deposited in the National Center for Biotechnology Information Sequence Read Archive.

Fungal Community Analysis

All statistical analyses were performed in the R environment (R Core Development Team 2012). Fungal sequences from the pooled samples were assigned to individual samples (i.e. demultiplexed) using Illumina bcl2fastq pipeline (v.2.20.0), and sequencing adapters were removed from the fungal ITS sequences using cutadapt (v.1.15)(Martin, 2011). Illumina-sequenced amplicon data of microbial communities is often constructed into operational taxonomic units (OTUs), in which sequences are clustered based on a fixed dissimilarity threshold. This clustering reduces the rate at which sequencing errors are misinterpreted as biological variation. The DADA2 package in R models and corrects Illumina-sequenced

amplicon errors and infers exact amplicon sequence variants (herein referred to as taxa), meaning variants detected due to biological variation and not due to sequencing noise (Callahan et al., 2016). This method can resolve biological differences at a high resolution, and the output can be directly compared between studies without the need to reprocess the pooled data. We employ DADA2 in this study. Specifically, quality control of sequencing reads for each sample consisted of truncating the first quality score of 2 (a quality score of 2 indicates a portion of the sequence that contains mostly low-quality reads of Q15 or less), and removing any with ambiguous base calls or higher than two expected errors. Reads shorter than 50 bases after quality trimming were removed.

Diversity

To compare the diversity of fungal communities among asymptomatic and symptomatic leaves of tall fescue, we quantified Hill's series of diversity. Diversity indicators are differentially sensitive to sample size and the abundance of rare and abundant taxa. Hill's series of diversity (Hill, 1973) is comprised of different orders (q) of diversity that summarize information about the number and relative abundances of taxa; diversity indices with different values of q are distinguished by the weighting applied to taxa that differ in abundance (Bent & Forney, 2008). We estimated taxa richness (Hills' N0, q=0), Shannon entropy (antilogarithm of the Shannon diversity, Hill's N1, q=1), and inverse Simpson diversity (Hill's N2, q=2) (Jost, 2006). Because Shannon entropy and Simpson diversity are less sensitive to the detection of rare taxa than species richness, they each place more weight on abundant taxa. Simpson diversity gives more weight to common or dominant taxa than Shannon diversity. The unit of each of the numbers within Hill's series of diversity is effective number of taxa, allowing comparisons across each value of diversity, and facilitate investigation of patterns of diversity by varying how

abundance is incorporated into each metric, Hill's numbers were calculated in R v. 3.5.0 (R Core Team 2013) with the vegan package (version 2.5.3)(Oksanen et al. 2013).

To test whether fungal diversity is associated with parasite symptoms, we used linear mixed models to explain Hill's N0, N1 and N2. We included the leaf segment characteristics (7 levels: Fully Asymptomatic, and the asymptomatic and symptomatic segments from leaves with each of the three focal parasites) and the square root of the number of sequencing reads obtained for a sample as fixed effects and leaf ID nested within collection site as random effects. We log-transformed the diversity indices to meet the normality assumption. High-throughput sequencing of pooled samples can result in samples that differ in sequencing depth. We accounted for observational bias stemming from this by incorporating this predictor, hereafter referred to as sequencing depth, into the models as a fixed effect, following Balint et al., 2015 (Figure 5.2). Linear mixed effects models were assessed in nlme, and we used emmeans (version 1.3.2) to evaluate the estimated marginal means of explanatory variables in explaining diversity variation, adjusted for multiple comparisons (Tukey HSD). We compared the partial residuals of Hill's numbers among the treatments with Tukey's HSD, after accounting for the variation caused by differential sequencing depth.

In the case of leaf segments symptomatic of *R. solani,* we took advantage of a 5-locus phylogeny of *Ceratobasidiaceae*, to investigate fungal diversity patterns excluding the taxa that may be causing the symptom (Gónzalez et al., 2016). *Ceratobasidiaceae* is a monophyletic group comprised of the genera *Ceratobasidium* and *Thanatephorus*, the teleomorphs of anamorphic *Rhizoctonia* fungi. While these fungi have been documented as saprobes and beneficial endomycorrhizal symbionts of orchids (Jiang, Lee, Cubeta, & Chen, 2015), they are parasites on grasses (L. Burpee & Martln, 1992). Using phylogenetic placement through T-BAS

(Carbone et al., 2019), we identified taxa that placed within the family (Gónzalez et al., 2016). We then created datasets in which these taxa were excluded and repeated the diversity analyses described above.

Since severity is a measure of disease progression, we used severity as a predictor to investigate if and how fungal diversity changed as disease progresses. Disease severity was assessed visually by estimating the percent area of a given leaf damaged by the parasite of interest. We investigated if and how diversity estimates correlated with *Rhizoctonia* severity among the *Rhizoctonia* symptomatic samples, if and how diversity estimates correlated with *Colletotrichum* severity among the *Colletotrichum* symptomatic samples, and if and how diversity estimates correlated with *Puccinia* severity among the *Puccinia* symptomatic samples. Specifically, we used linear models with square root of sequencing read numbers obtained for a sample and the leaf segment characteristics (asymptomatic/symptomatic; parasite) as explanatory variables. Each model of the numbers in Hill's series included sequencing depth and disease severity as fixed effects and site as a random effect (Hill ~ sqrt(readNumbers) + RhizDamage + 1|Site). A significant effect of disease severity on fungal diversity associated with a leaf segment would be consistent with the parasite gradually impacting the fungal microbiome as the disease progresses. No relationship between disease severity and fungal diversity associated with a leaf segment would suggest that any association between the parasite and microbiome is not due to a gradual impact of the parasite on the microbiome.

Community Composition

To test the hypothesis that fungal taxa composition may be different when symptoms of parasites that kill tissue are observed, we tested whether fungal taxa composition at the leaf segment level was correlated with disease properties. Bray-Curtis distances were calculated

among samples separately. Dissimilarities were visualized using non-metric multidimensional scaling (NMDS) implemented in the phyloseq package. We performed permutational MANOVAs using the adonis function in vegan. Predictors included Site and Parasite Status. The adonis function is sensitive to the order in which variables are added, so models were run multiple times, varying the order of predictors, to verify important predictors and we report predictors that were significant regardless of order.

Representative sequence data representing each taxon identified by DADA2 was compared against UNITE via BLAST. Additionally, sequences were evaluated for phylogenetic placement in T-BAS applying maximum likelihood with RAxML and default settings to place query sequences into a reference fungal phylogeny using the ITS dataset (Carbone et al., 2017).

We also assessed if the average abundance of particular fungal groups in a sample varied with symptom, using glmFit and glmLRT in the package edgeR following McMurdie and Holmes (2014). We included taxa that occurred within 5% of samples included in these analyses. Following the manual for edgeR, we performed a paired test in which we investigated the differences due to symptom after accounting for baseline variation between leaves (in the case of comparisons between the asymptomatic and symptomatic segments of infected leaves) by using an additive model formula that included leaf as a predictor. We used glmFit to fit a negative binomial generalized log-linear model to abundance of each taxon, and we used glmLRT to conduct taxon-wise statistical tests for a contrast between parasite symptom categories. Following McMurdie & Holmes, 2014, all tests were corrected for multiple inferences using the Benjamini-Hochberg method to control the false discovery rate (Benjamini & Hochberg, 1995). We used a false discovery rate (FDR) cutoff of 0.05.

Results

From the 252 leaf segments, Illumina generated 6,650,600 ITS1 reads. Of these, 4,483,694 read pairs passed quality filtering. This represents a mean number of reads per sample of 17,792. Using DADA2, we identified 2961 unique amplicon sequencing variants (taxa). This represents a mean number of taxa per sample of 70.8 (median of 62). Using T-BAS, all taxa placed within the kingdom fungi. 12.5% of taxa could not be placed lower than the kingdom fungi. Of the taxa that could be placed lower than the kingdom fungi, 99.2% placed within Ascomycota (1459) or Basidiomycota (1110). At the class level, most taxa within Ascomycota were assigned to Dothideomycetes (808) or Sordariomycetes (307), and most taxa within Basidiomycota were assigned to Agaricomycetes (691) or Tremellomycetes (186). The following analyses consider these 2961 taxa delineated by DADA2.

Diversity

After accounting for variation in sequencing depth, parasite symptom category strongly and significantly predicted variation in all three numbers in Hill's series (ANOVA $p < 0.0001$). There were fewer fungal taxa (Hill's N0) and there was lower diversity (Hill's N1 and N2) in leaf segments that exhibited symptoms of *Rhizoctonia solani* compared to leaf segments that were either asymptomatic or symptomatic of other fungal parasites (Table 5.2, Figure 5.3, C5.1, Tukey's HSD, $p < 0.01$). Specifically, when comparing the mean of each of Hill's numbers between leaf segments symptomatic of *R. solani* and leaf segments within all other parasite symptom categories, Hill's N0 was 41.4-55.0% lower, Hill's N1 was 66.3-77.6% lower, and Hill's N2 was 58.8-71.1% lower. We collected leaves that exhibited symptoms of a parasite and from each leaf, cut two leaf segments: a segment in which the symptom was observed and a segment that was asymptomatic. Leaf segments exhibiting symptoms of *R. solani* had lower

fungal richness (by 49.3%) and diversity (Hill's N1 by 72.4% and Hill's N2 by 64.1%) on average when compared to the asymptomatic segments taken from the same leaves (Tukey's HSD, $p < 0.001$). This suggests that the fungal diversity of the leaf as a whole was not lower prior to invasion by *R. solani*.

To investigate the lower richness of samples symptomatic of *R. solani,* we considered if and how the diversity metrics as defined in Hill's series varied by estimated disease severity (estimated percent leaf area damaged by a given parasite) within just the segments symptomatic of *R. solani*. Fungal diversity generally increased with *R. solani* disease severity, with increasing statistical support for parameters that give less weight to relative abundance. (Table 5.3, Figure 5.4, Hill's N0 $p = 0.010$, Hill's N1 $p = 0.043$, Hill's N2 $p = 0.064$). This suggests that fungal diversity, particularly richness, does not decrease progressively as *R. solani* spreads through the leaf.

We identified 26 taxa that placed within the family *Ceratobasidiaceae*. We investigated if and how fungal diversity was associated with parasite symptom category when these taxa, which could be the cause of the symptom, are excluded from analyses. After accounting for differences in sequencing depth, parasite symptom category strongly and significant predicted variation in all three numbers in Hill's series ($p < 0.001$, Table 5.4, Figure 5.5). Across the three numbers in Hill's series, fungal diversity was lowest within leaf segments that were symptomatic of *R. solani*, though this lower fungal diversity did not have statistical support when compared to that of segments that were asymptomatic or asymptomatic and elsewhere infected with *P. coronata* (Tukey's HSD $p > 0.05$). Specifically, when comparing the mean of each of Hill's numbers between leaf segments symptomatic of *R. solani* and leaf segments within all other parasite symptom categories, Hill's N0 was 30.3-50.1% lower, Hill's N1 was 54.1-69.1% lower,

and Hill's N2 was 43.7-62.0% lower. Within leaf segments symptomatic of *R. solani,* fungal diversity still had a significant correlation with estimated disease severity, even when excluding those taxa placed within *Ceratobasidiaceae* (Figure 5.6, $p < 0.05$). The association between *R. solani* symptom and lower fungal diversity therefore holds when just considering the fungal components that are not the cause of the symptom.

Community Composition

The community composition of leaf segments with symptoms of *R. solani* differed from leaf segments that were asymptomatic or symptomatic of other parasites (Table 5.6, Figure 5.7, PerMANOVA, $p = 0.001$). Therefore, symptoms of *R. solani* are associated with different fungal taxa compared to the other categories of leaves. Within the subset of samples that exhibited symptoms of *R. solani*, estimated disease severity was also a significant predictor of taxa composition, but only accounted for a modest amount of variation in fungal taxa composition (Figure 5.8, PerMANOVA, $R^2=0.08$, $p = 0.043$).

Similar to diversity analyses, we repeated analyses of fungal taxa composition on a subset of taxa that excluded those that placed within *Ceratobasidiaceae*. The taxonomic composition of leaf segments with symptoms of *R. solani* still differed from leaf segments that were asymptomatic or symptomatic of other parasites (Table 5.5, Figure 5.9, PerMANOVA, $p = 0.001$). Within the subset of samples that exhibited symptoms of *R. solani*, estimated disease severity was still a significant predictor of taxa composition and still only accounted for a modest amount of variation in fungal taxa composition (Figure 5.10, PerMANOVA, $R^2=0.10$, $p = 0.02$). Therefore, the difference in composition was not only due to the presence of taxa that were causing the symptom.

We investigated how relative abundance of the 2961 taxa identified varied between leaf segments symptomatic of *R. solani* and the asymptomatic segments that were infected with *R. solani* elsewhere. After accounting for baseline differences among leaves by including a random effect for leaf, we identified 11 taxa, representing ten genera, that differed significantly ($p < 0.05$) between these two types of leaf segments (Figure 5.11). Leaf segments symptomatic of *R. solani* had higher relative abundance of a taxon that placed within *Ceratobasidium* than the asymptomatic segments (log fold change of 9.40 (variant 153). Other taxa that had higher relative abundance in symptomatic than asymptomatic tissue placed within *Ramichloridium* (2 taxa: Taxon 185: log fold change of 7.57; Taxon 23: 10.00), *Glomerobolus* (log fold change of 8.42, *Wojnowicia* (log fold change of 6.64), *Nigrospora* (log fold change of 6.43), *Pyrenochaetopsis* (log fold change of 6.36), and *Cryptococcus* (log fold change of 6.22). Leaf segments symptomatic of *R. solani* had lower relative abundance of three taxa: one within *Psilocybe* (log fold change of 4.31), one within *Phaeosphaeria* (log fold change of 4.55) and one within *Apenidiella* (log fold change of 7.77).

Discussion

This study aimed to evaluate how the fungal microbiome associates with symptoms of varying and co-occurring parasites under field conditions. While symptoms of one parasite were associated with less diverse and compositionally distinct host-associated fungal microbiota compared to segments that were asymptomatic of this parasite, symptoms of the other two parasites were not associated with such differences. We highlight the need to consider parasite species traits in order to better understand parasite-specific associations between parasites and the microbiome, though replication of parasite traits would be necessary to definitively implicate certain mechanisms.

The fungal communities of leaf segments symptomatic of *R. solani* were less diverse and distinct from those of segments that were asymptomatic or symptomatic of *C. cereale* or *P. coronata*. These results are consistent with findings in which host microbial communities are less diverse with parasite infection (Leung et al., 2018; Wu et al., 2014). In contrast, fungal diversity and composition of leaf segments symptomatic of *C. cereale* and *P. coronata* were not significantly different from those of asymptomatic segments. These results are consistent with findings in which there is no association between parasite infection and host-associated microbiota (Baxter et al., 2015; Kreisinger, Bastien, Hauffe, Marchesi, & Perkins, 2015). *R. solani*, a necrotroph, extracts resources by killing host cells, while *P. coronata*, a biotroph, extracts resources from live host cells. We predicted that *R. solani* may act as a niche modifier, changing the host-associated fungal community as it creates necrotic tissue while it spreads. Because *P. coronata* does not modify the host environment in the same way, we predicted it would not be associated with such a change. Our results support these predictions and the lower fungal diversity and richness associated with *R. solani* suggest that the parasite may make the host environment less suitable for fungal taxa that require live host tissue.

C. cereale infects host tissue in two stages: an initial asymptomatic stage in which it extracts resources from live host cells, followed by a symptomatic stage in which it kills host tissue to extract resources. Given that symptoms of *C. cereale* are indicative that the fungus is in its second, necrotrophic stage of infection, we expected that there may be a difference in the fungal community associated with a leaf symptomatic of *C. cereale* compared to an asymptomatic leaf. However, our results showed that the diversity and composition of leaf segments exhibiting symptoms of this parasite were no different from segments asymptomatic of

the parasite. Future work that disentangles the complex disease cycle of *C. cereale* and its association with the microbiome at each stage of its infection process may elucidate these results.

We propose niche modification, as determined by parasite feeding strategy, as a mechanism to explain the correlation between damage from *R. solani* and a less diverse fungal microbiome. Specifically, as *R. solani* may modify the host environment by creating necrotic tissue, thereby impacting the diversity and composition of fungi that can inhabit that environment. Similar patterns were also seen in a study that documented the relationship between host microbiome and another parasite that kills host cells while it spreads; specifically, an *Eimeria* parasite, which colonizes the vertebrate gut and kills epithelial cells as a consequence of the intracellular stages of its life cycle, was associated with lower bacterial richness in its host (Wu et al., 2014). Our results indicated that disease severity, as estimated by percent leaf damaged, of *R. solani* had a positive correlation with fungal richness and diversity. These results suggest that the fungal community does not gradually become less diverse as the parasite spreads, as that would be suggested by a negative relationship between disease severity and fungal diversity. The initial invasion of *R. solani* on host tissue may disrupt the microbiome, as the parasite makes the tissue less suitable for fungi that rely on live host tissue. As the parasite spreads, the necrotic tissue may be colonized by saprobes. Because we collected samples from a naturally-occurring epidemic, we were limited to naturally-occurring infections. Most natural infections had relatively low disease severity, which limited our ability to test a full range of disease severity. Considering the associations between the microbiome and different parasites will be necessary to draw a definitive link between parasite feeding strategy and effects on the microbiome. Studies of other necrotrophic and biotrophic parasite species that similarly associate with the host microbiome would be consistent with our hypothesis. Experimental manipulations

in which the microbiomes of host individuals are sampled longitudinally before and after parasite infection would be needed to provide evidence that is not just correlative.

As the importance of microbial communities to host health has been realized and continues to be interrogated, it is critical that the characterization of these communities is accurate and efficient. Microbiome studies have historically relied on constructing operational taxonomic units, or clusters of sequences that differ by a given dissimilarity threshold, to disentangle biological variation from sequencing error (Schloss & Handelsman, 2005; Sun et al., 2009). By clustering sequences, the ability to discern fine-scale variation is lost (Eren, Borisy, Huse, & Mark Welch, 2014; Rosen, Callahan, Fisher, & Holmes, 2012), which can be informative about ecological and evolutionary processes. Additionally, dissimilarity thresholds vary by study, which can limit comparisons across studies. Here, we utilized DADA2 to model and correct Illumina-sequenced amplicon error within our sequencing (Callahan et al., 2016). This method fully utilizes sequence-abundance information. Critically, by inferring true sequence variants without setting study-specific dissimilarity thresholds, comparisons across studies will be easier. As the field moves to understand the mechanisms that explain parasite species-specific associations with the microbiome, the ability to readily compare across studies will be essential to discern any generalities.

Recent work by Leung et al. 2018 clearly demonstrated that associations between parasites and host microbiome are strongly affected by environmental factors, and that effects of microbial interactions under lab settings may not scale up to those under field conditions (Leung et al., 2018). It will therefore be critical to conduct more inquiries of parasite-microbiome interactions under field settings. Foliar fungal parasites have been studied extensively (Busby, Peay, & Newcombe, 2016; Halliday, Heckman, Wilfahrt, & Mitchell, 2017; Spear & Mordecai,

2018) and can serve as a suitable model system to investigate microbiome/parasite interactions under field settings. Not only is their consequent disease of interest, they are often readily identifiable by symptom, which facilitates observational and longitudinal studies in the field. Additionally, plant fungal parasites are diverse and vary in reliance on a host, host specificity, feeding strategy, and stimulated host immune response, so may especially cater to studies considering variation among parasite species in parasite-microbiome studies.

In summary, our results demonstrate that the diversity and composition of the fungal microbiome can associate with parasites, but such associations are specific to parasite species. Although the mechanisms responsible for the patterns we observed remain unclear, our analyses indicated that damage from *R. solani* is associated with fungal communities that are less diverse and compositionally distinct. While we put forward that *R. solani* may act as a niche modifier, making the host environment less suitable for certain fungal groups, manipulative experiments would be needed to clearly determine if the parasite is affecting the microbiome. We suggest that a key to understanding parasite-specific associations with host-associated microbiota may be considering parasite species traits, like parasite feeding strategy and stimulated host immune response, particularly if and how a parasite changes its host environment.

Table 5.1 Biology and ecology of the focal fungal parasites

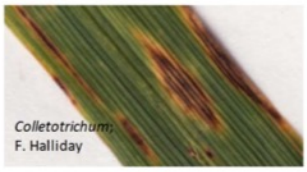
Parasite	Feeding Strategy	Representative symptom
Rhizoctonia solani	**Necrotroph**: kills the living cells of a host and then feeds on the dead matter	 *Rhizoctonia*; F. Halliday
Puccinia coronata	**Biotroph**: feeds on living cells of a host, without killing it as part of the infection process	 *Puccinia*; K. O'Keeffe
Colletotrichum cereale	**Hemibiotroph**: initially feeds on living host tissue without causing visible symptoms prior to switching to a necrotrophic phase	 *Colletotrichum*; F. Halliday

Table 5.2: Variation of Hill's series of diversity, explained by linear mixed models. Hill's N0 is species richness; N1 is the antilogarithm of Shannon's diversity, and N2 is the inverse of Simpson's Diversity

	Hill's N0					Hill's N1					Hill's N2			
	F	denDF	numDF	P		F	denDF	numDF	P		F	denDF	numDF	P
Fixed Effects														
Read numbers	69.626	84	1	<.0001		6.342	84	1	0.004		5.032	84	1	0.0275
Parasite Symptom	6.908	84	6	<.0001		10.994	84	6	<.0001		9.317	84	6	<.0001

Table 5.3: Leaf-associated fungal diversity and richness has positive correlation with estimated disease severity (percent leaf damaged).

	Hill's N0					Hill's N1					Hill's N2			
	F	denDF	numDF	P		F	denDF	numDF	P		F	denDF	numDF	P
Fixed Effects														
Read numbers	9.872	35	1	0.004		1.973	84	1	0.171		1.390	35	1	0.2483
Rhiz. Damage	7.642	35	1	0.010		4.500	84	1	0.043		3.723	35	1	0.0639

Table 5.4: Variation of Hill's series of diversity, explained by linear mixed models, for the subset of variants that do not place within Ceratobasidiaceae. Hill's N0 is species richness; N1 is the antilogarithm of Shannon's diversity, and N2 is the inverse of Simpson's Diversity

	Hill's N0					Hill's N1					Hill's N2			
	F	**denDF**	**numDF**	**P**		**F**	**denDF**	**numDF**	**P**		**F**	**denDF**	**numDF**	**P**
Fixed Effects														
Read numbers	87.686	84	1	<.0001		5.917	84	1	0.0171		1.735	84	1	0.1914
Parasite Symptom	6.259	84	6	0.002		5.634	84	6	0.001		4.810	84	6	0.0003

Table 5.5: Variation in the composition of the fungal microbiome explained by parasite and symptom. Results of adonis (permutational MANOVA). Predictors were included in the order indicated below. However, relative importance remained unchanged among different permutations. Values indicate the R^2

	denDF	**numDF**	**R^2**	**p**
Fixed Effects				
Read numbers	251	1	0.008	0.011
Parasite Symptom	246	6	0.101	0.001

Figure 5.1. Map of the transects in the study. Mapped area is a grass-dominated field that is part of the Duke Forest Teaching and Research laboratory in Orange County, North Carolina. Green lines indicate the approximate location of each transect across the field, and each transect was at least 20 meters apart. The inset indicates how each transect was divided into sampling sites. Samples were collected every 20 meters along a 100 meter transect.

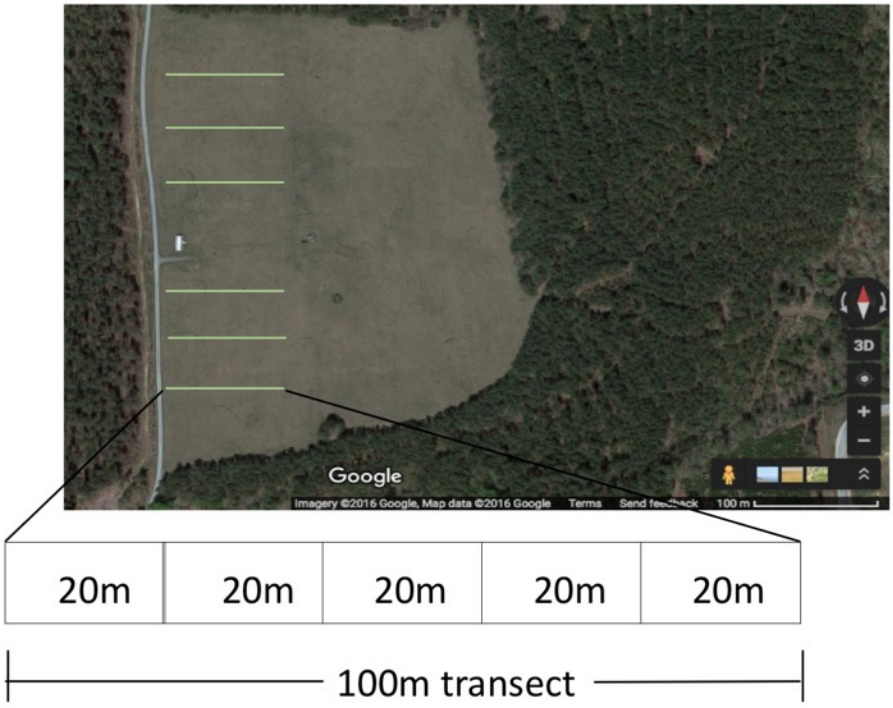


Figure 5.2. Hypothesized effects of sequencing depth and parasite symptom on fungal diversity. A. Sequencing depth results can result from technical rather than biological reasons. When multiple samples are sequencing in the same reaction, the resulting read numbers can vary by orders of magnitude (Schmidt et al. 2013, Balint et al 2015). Relatively higher sequencing depth could lead to a higher number of taxa identified in a given sample, resulting in a misinterpretation of that sample having a relatively higher fungal diversity. Such a misclassification would be strongest Hill's N0, which is simply the number of taxa in a sample and does not take evenness into account. B. Fungal diversity may be associated with the presence of certain fungal parasites. The fungal microbiome of a host may affect the host's susceptibility to and tolerance of a given parasite. Parasites may also disrupt the microbiome as they invade host tissue, which could lead to a change in fungal diversity. Parasites may increase fungal diversity, perhaps in the case of necrotrophic and hemibiotrophic parasites like *R. solani* and *C. cereale* in which plant tissue is ultimately killed during infection, creating niches for saprobes.

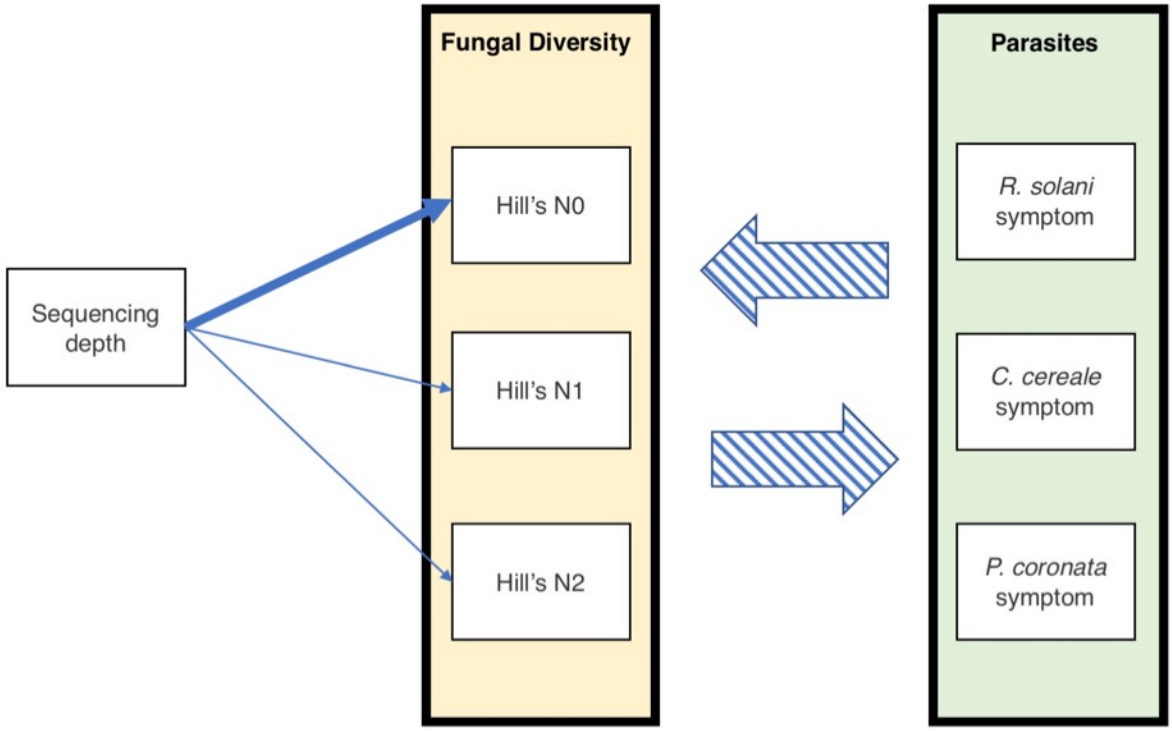


Figure 5.3 Leaf segments with symptoms of *Rhizoctonia solani* were associated with less diverse foliar fungal communities. Panels show taxa diversity estimated using Hill numbers which correspond to observed species richness (N0), exponent of Shannon entropy (N1), and (c) inverse Simpson's diversity (N2). Leaf segments with *Rhizoctonia* lesions had statistically significantly lower diversity in comparison to the other treatments, evaluated with Tukey's HSD. Letters mark statistically significant differences in Hill's numbers evaluated with Tukey's HSD. The shown estimated marginal means ± 95% confidence intervals.

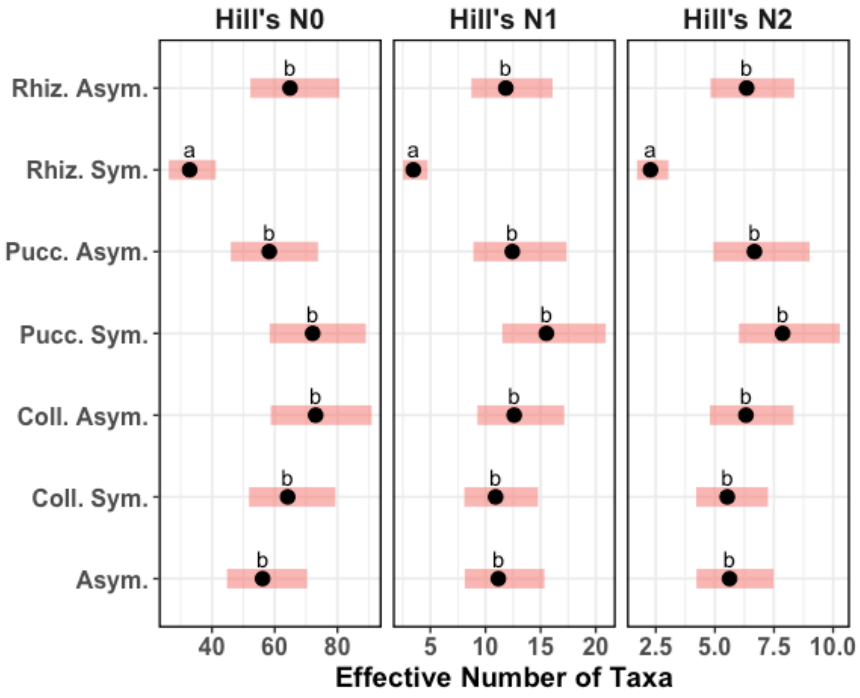


Figure 5.4. Leaf-associated fungal diversity and richness has positive correlation with estimated disease severity (percent leaf damaged). Panels show species diversity estimated using Hill numbers which correspond to observed species richness (N0), exponent of Shannon entropy (N1), and (c) inverse Simpson's diversity (N2). Points represent diversity estimated from the ITS region (Illumina sequencing). Lines represent best-fit regressions between disease severity and the diversity metric. While Hill's N0 had a significant relationship with *R. solani* disease severity ($p = 0.0100$), there was no statistically significant relationship between *R. solani* disease severity and Hill's N1 ($p = 0.0429$) or Hill's N2 ($p = 0.0639$).

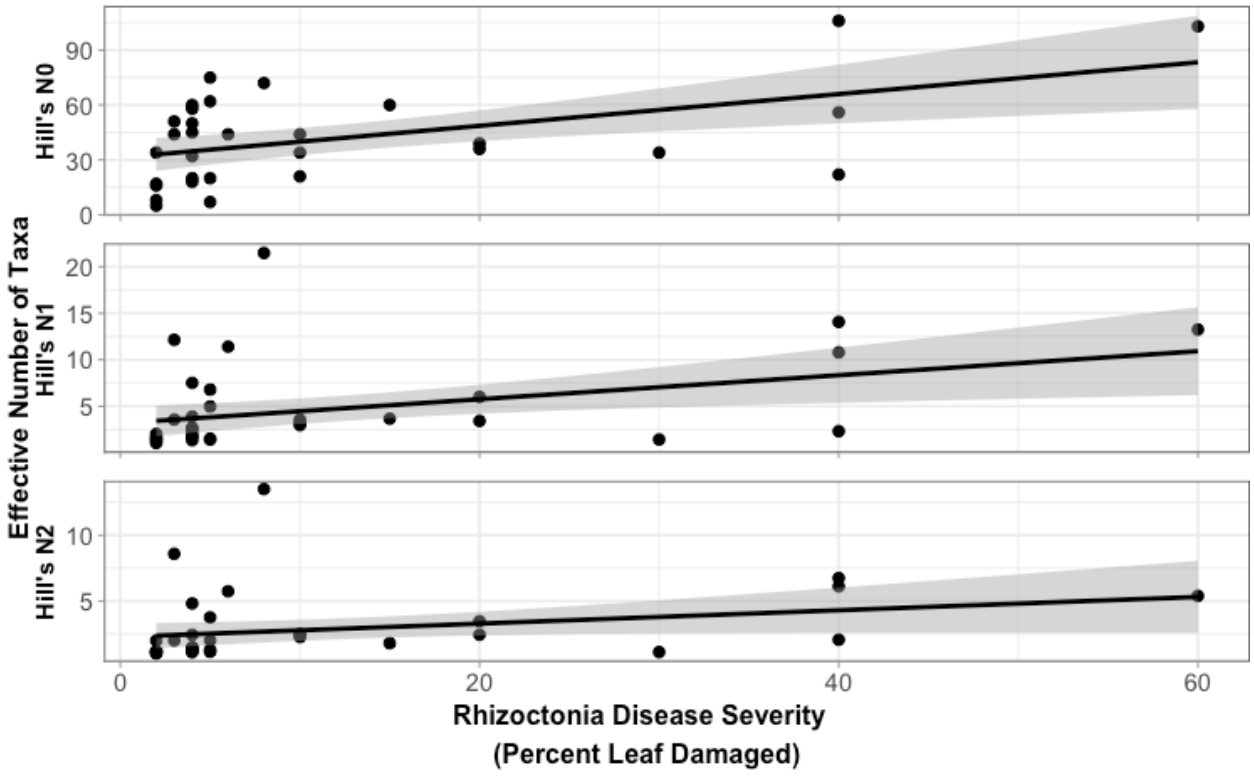


Figure 5.5 Leaf segments with symptoms of *Rhizoctonia solani* were generally associated with less diverse foliar fungal communities of taxa that did not place within *Ceratobasidiaceae*. Smear plots show the Hill diversity profile of the foliar microbiome among healthy leaves, asymptomatic and symptomatic segments from leaves infected with each of three focal parasites. Means ± 95% confidence intervals are shown are from the estimated marginal means of our linear mixed model. Diversity differences are consistent across the Hill diversity profile. Panels show species diversity estimated using Hill numbers which correspond to observed species richness (N0), exponent of Shannon entropy (N1), and (c) inverse Simpson's diversity (N2). Letter above each point indicate statistical differences ($p < 0.05$).

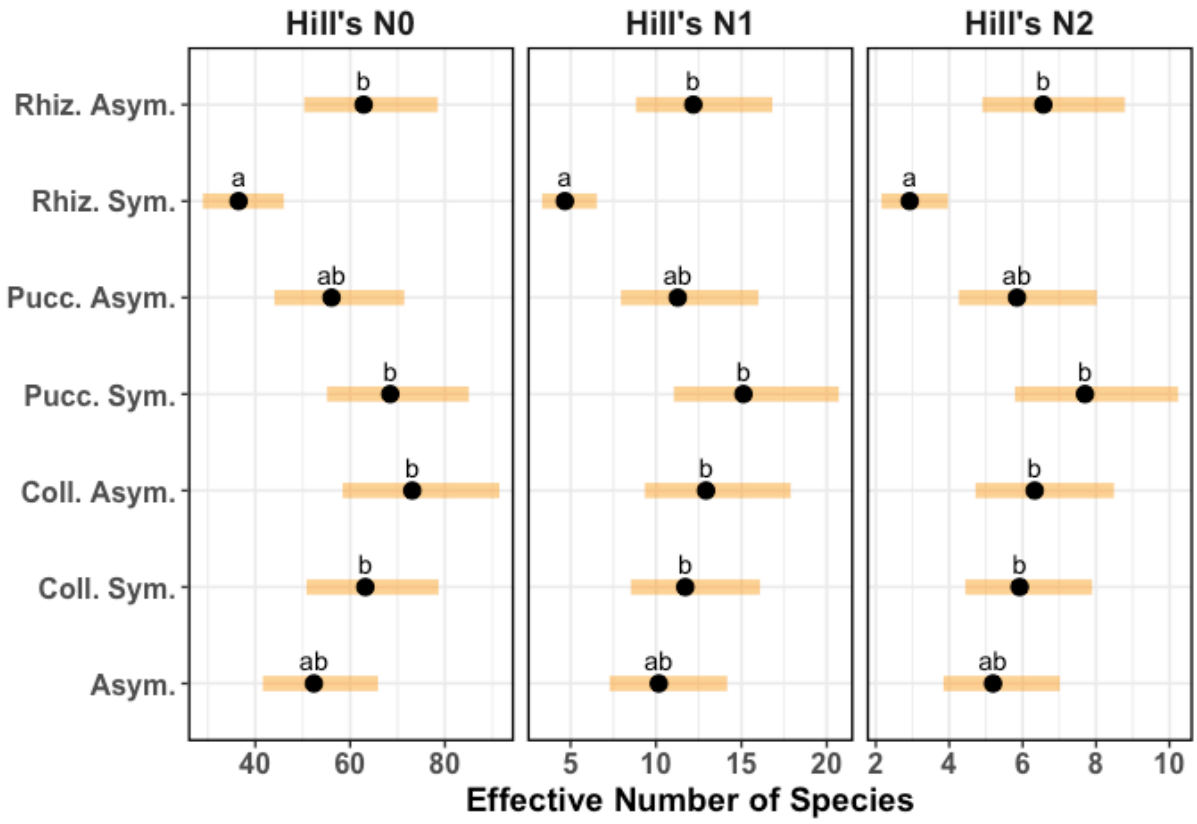


Figure 5.6. After excluding taxa placed within *Ceratobasidiaceae,* leaf-associated fungal diversity and richness has positive correlation with estimated disease severity (percent leaf damaged). Diversity was estimated at the leaf segment level and regressed against estimated disease severity (area leaf damaged). Panels show species diversity estimated using Hill numbers which correspond to observed species richness (N0), exponent of Shannon entropy (N1), and (c) inverse Simpson's diversity (N2). Points represent diversity estimated from the ITS region (Illumina sequencing). Lines represent best-fit regressions between disease severity and the diversity metric. All of Hill's numbers had a significant relationship with *R. solani* disease severity ($p < 0.05$).

Figure 5.7. Non-metric multidimensional scaling (NMDS) plots showing similarity among samples based on fungal taxa composition. Sequencing generated using Illumina sequencing of the ITS region, and composition based on the Bray-Curtis distance metric. Points represent individual leaf segments. Color indicates the parasite (*Rhizoctonia*, *Colletotrichum*, or *Puccinia*) and symptom status of a given leaf segment.

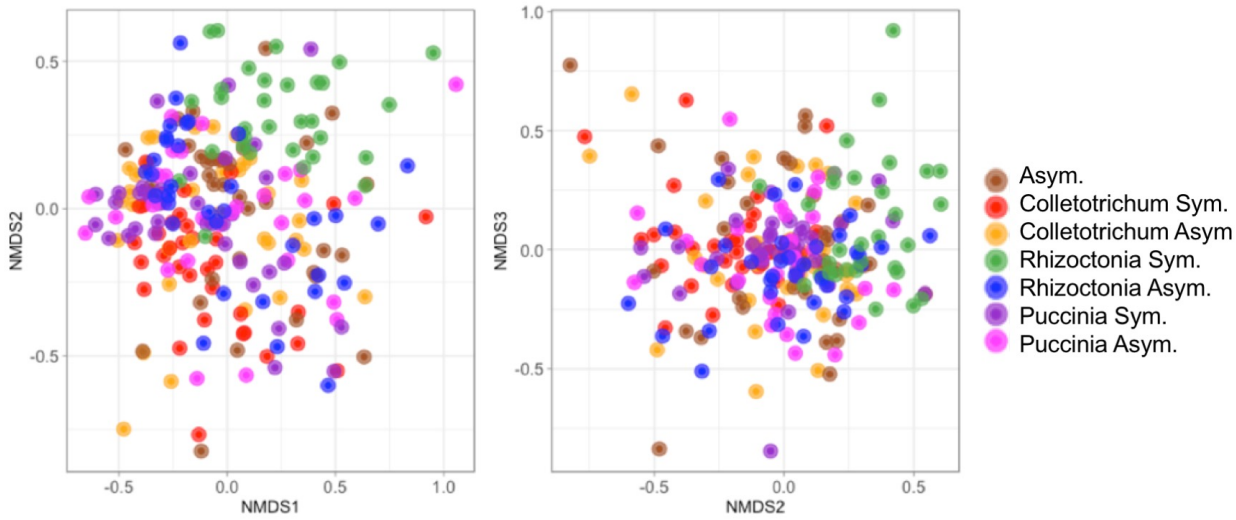


Figure 5.8. Estimated severity of damage caused by *R. solani* significantly predicted fungal taxa composition, but only explained a modest amount of variation Nonmetric multidimensional scaling (NMDS) plots showing similarity among samples that exhibited *Rhizoctonia* symptoms based on fungal taxa composition generated using Illumina sequencing of the ITS region, based on the Bray-Curtis distance metric. Points represent individual leaf segments. Point size indicates estimated disease severity as determined by percent leaf damaged by *Rhizoctonia*.

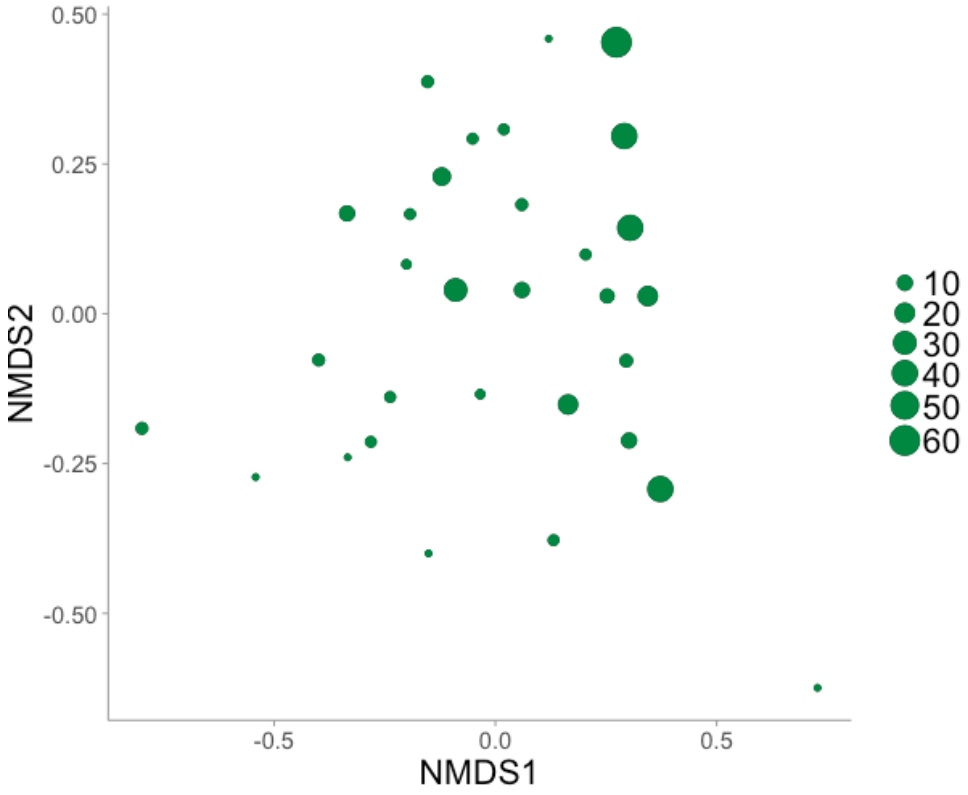


Figure 5.9. Non-metric multidimensional scaling (NMDS) plots showing similarity among samples based on composition of fungal taxa excluding those taxa that placed within *Ceratobasidiaceae* generated using Illumina sequencing of the ITS region, based on the Brays-Curtis distance metric. Points represent individual leaf segments. Color indicates the symptom category of a given leaf segment.

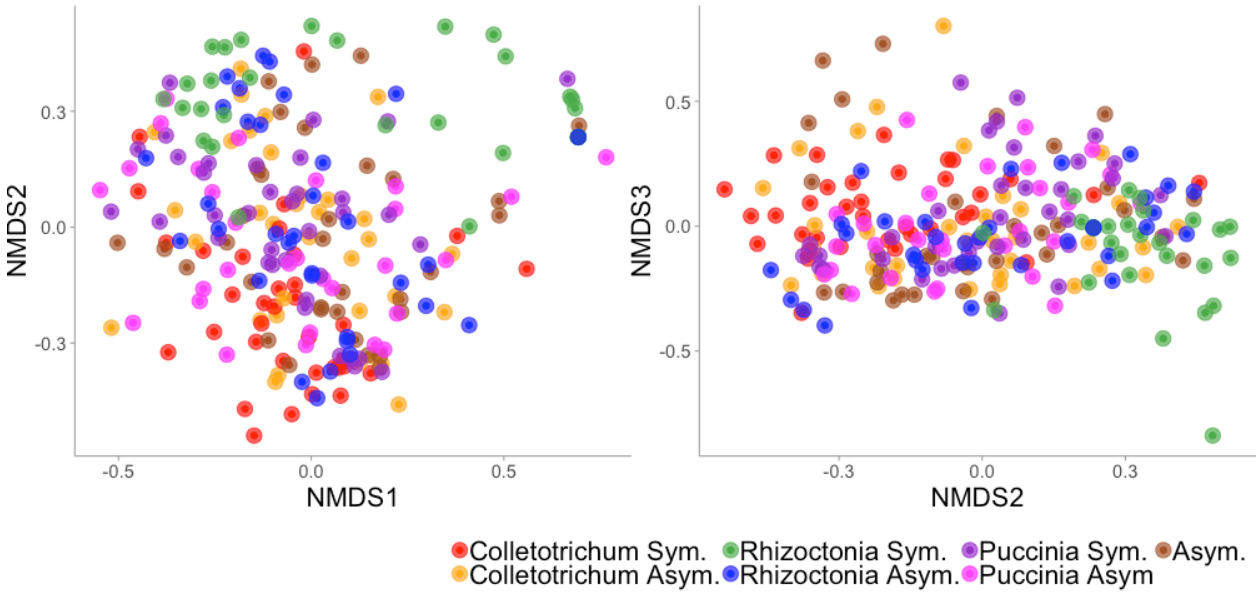


Figure 5.10. Estimated severity of damage caused by *R. solani* significantly predicted fungal taxa composition excluding those taxa that placed within *Ceratobasidiaceae*, but only explained a modest amount of variation. Nonmetric multidimensional scaling (NMDS) plots showing similarity based on fungal species composition generated using Illumina sequencing of the ITS region among samples that exhibited *Rhizoctonia* symptoms, based on the Bray-Curtis distance metric. Points represent individual leaf segments. Point size indicates estimated disease severity as determined by percent leaf damaged by *Rhizoctonia*.

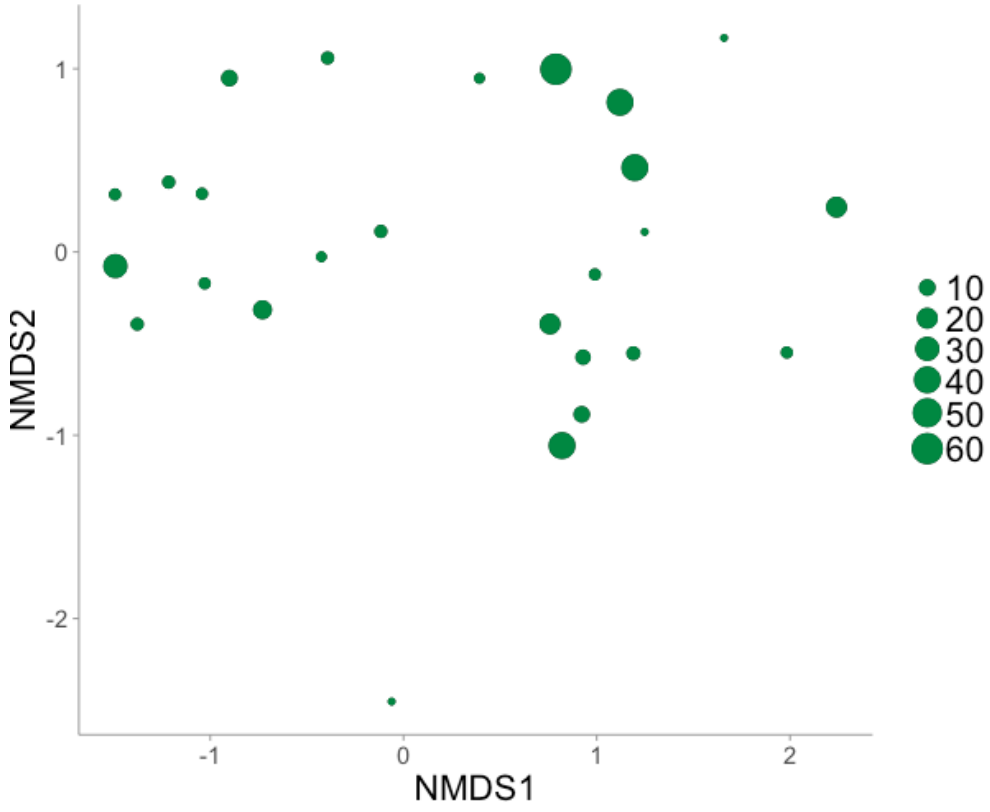
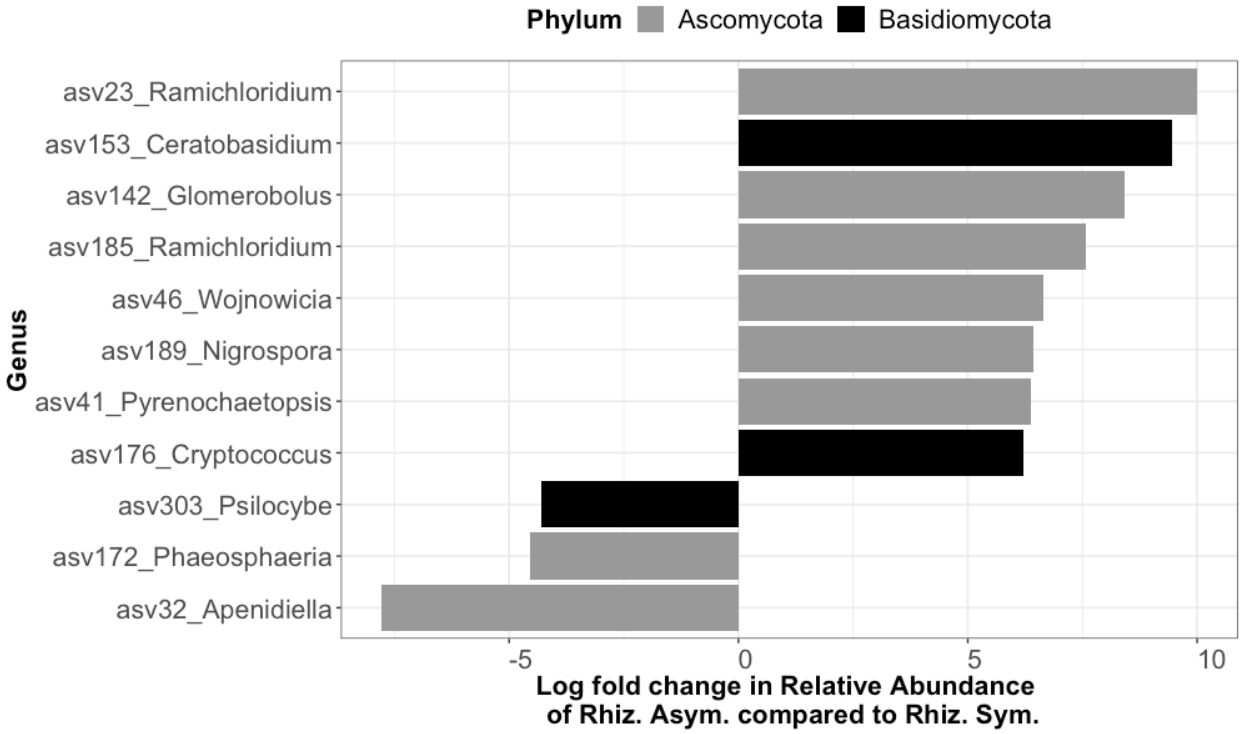


Figure 5.11: 11 taxa, representing 10 genera, differed significantly in relative abundance between the asymptomatic and symptomatic segments of leaves infected with *R. solani*. Results of edgeR analysis comparing relative sequence abundance for fungal sequencing taxa between the asymptomatic and symptomatic segments of leaves infected with *R. solani*. Bars represent the log fold difference in relative sequence abundance for taxa that differed significantly between asymptomatic and symptomatic segments ($p < 0.05$) with the genus name indicated on the x-axis.



CHAPTER 6: CONCLUSION

It has become clear that coinfection is common in plant and animal hosts (Bordes & Morand, 2011; Tollenaere et al., 2016). Incorporating increasing levels of complexity, from a one-host/one-parasite framework to including multiple parasite species, has enhanced our understanding of the implications of coinfection on disease dynamics. However, most of these insights into the implications of coinfection on disease dynamics are from lab studies of pairwise interactions between symbionts (Adame-Álvarez et al., 2014; Graham, 2008; Pańka et al., 2013), which may neglect occurrences in which interactions take place among more than two symbionts (Telfer et al., 2010). Considering higher-order interactions, indirect interactions among species through effects on intermediate species (Wootton, 1994), could certainly enhance our understanding of the ecological consequences of coinfections (Marchetto & Power, 2018). Further, most studies consider impacts of coinfections within host individuals, but such impacts can have effects far beyond the individual (O'Keeffe et al., 2017). The research presented in this dissertation aimed to fill this need by integrating studies within one system that consider coinfections within the host individual, host population, and as part of the host-associated fungal community to better understand what the impacts of within-host microbial interactions are for a fungal plant parasite.

Microbial interactions within host individuals can have impacts that shape disease in the long-term. Within-host microbial interactions can affect parasite transmission rates, spatial and temporal patterns of parasite abundances and parasite co-occurrence (Al-Naimi, Garrett, &

Bockus, 2005; P. A. Clay et al., 2018; Halliday, Umbanhowar, et al., 2017). Chapter Four found that populations inoculated with a systemic endophyte did not experience a change in disease intensity caused by a parasite over time but did experience a higher peak prevalence of that parasite. The mutualistic relationship between *Epichloë* endophytes and cool-season grasses has been studied extensively, but this experiment, in which *E. coenophiala* had no or a positive effect on parasite, *R. solani*, is consistent with evidence that the systemic endophyte does not always provide protection against parasites (L. L. Burpee & Bouton, 1993; Halliday, Umbanhowar, et al., 2017; Wäli et al., 2006). Importantly, we were able to detect the consequences of this interaction under field conditions, and field mesocosm experiments may offer the opportunity to gain experimental evidence on the impacts of coinfection on parasite transmission rates.

Exploring this interaction further, Chapter three documented interactions among these two fungal symbionts, as well as another fungal parasite, *C. cereale*, within a shared host under controlled settings. Within a host individual, interactions between parasites and coinfecting microbes can affect parasite growth, disease severity, and host mortality (Pieter T J Johnson & Hoverman, 2012; K. Lee, Pan, & May, 2009; Pinna et al., 2016). In the experiments described in Chapter three, one parasite, *C. cereale,* facilitated the growth of the other parasite, *R. solani*. The magnitude of this facilitation increased when the host was also infected with endophyte, *E. coenophiala*. These results suggest that higher-order interactions can play an important role in within-host parasite growth. Each of these fungal symbionts coinfect host individuals and co-occur within host populations of tall fescue in the field (Halliday et al., 2017, Mitchell Lab unpublished data). There is mounting evidence that effects of microbial interactions under lab settings may not scale up to those under field conditions (Leung et al., 2018). In fact, Halliday et al. , 2017 documented an antagonistic interaction between *C. cereale* and *R. solani*. It may be

that a limitation to directly extrapolating findings from lab experiments to field settings is the presence of higher-order interactions that are difficult to detect. Future work that considers higher-order interactions under field settings may clarify this connection.

To further explore if and how fungal parasite *R. solani* associated with the community of host-associated microbiota (as well as how symptoms of other parasites associated), Chapter five documented a fungal community sequencing survey as associated with parasite symptoms on tall fescue under field conditions. This survey illustrated that leaf segments exhibiting symptoms of *R. solani* had a lower fungal diversity and distinct fungal composition compared to asymptomatic leaf segments, while segments symptomatic of other parasites did not exhibit any differences from asymptomatic leaf segments. Our results are consistent with findings that the interaction between parasites and the microbiome may be parasite species-specific. In this system, unlike the other parasites, *R. solani* extracts resources from its host exclusively by killing host cells. The other parasites spend part or all of their disease cycle extracting resources from live host cells. We put forward the hypothesis that *R. solani* acts as a niche modifier, as defined by Fukami, 2015, by impacting the host environment in such a way as to make the host more or less suitable to new colonizers. Similar association studies among parasites of varying feeding strategies and host microbial community would strengthen the link between our results and niche modification. Further, consideration of parasite traits, such as feeding strategy, transmission mode, and life cycle, may elucidate the mechanisms underlying parasite species specific responses to within-host microbial interactions.

As coinfections and more complex interactions are considered in relation to disease dynamics, such as what was presented in Chapters three, four, and five, gene expression studies may elucidate mechanisms underlying such interactions. Dual RNA-seq, global gene expression

sequencing of parasite-infected host tissue, can enhance understanding of the genetic mechanisms underlying host/pathogen interactions (Alexander J Westermann, Barquist, & Vogel, 2017) and potentially be expanded to consider within-host microbial interactions. While next-generation sequencing has certainly become more accessible in the last few years, it is critical that we assess the accuracy of analytical approaches. Using the genomic resources of a related species to analyze sequencing of a species in which genomic resources are unavailable or limited is commonly utilized in free-living species. Chapter two found that this approach can lead to inaccurate read mapping, with reads from the host mismapping to the genome of species related to the pathogen. This could have severe consequences for biological interpretation if appropriate steps are not taken. By proposing a workflow to avoid such inaccuracies, this chapter expanded the use of this sequencing approach to non-model systems, as it was previously limited to model systems in which genomic resources are developed and readily available. Expansion of these methods may expand this sequencing application to interrogation of within-host microbial interactions.

This dissertation showed that the microbial community surrounding a parasite shapes both that pathogen's immediate phenotype, as shown in a change in parasite growth rate, and its evolutionary potential, as shown in effects on peak prevalence of the same parasite in a population (Chapter Four). Further, it highlighted potential complexities in these effects, as higher-order interactions may play an important role in the onset of disease caused by this parasite (Chapter three) and the diversity and composition of the microbial community can associate with symptoms of this parasite but not symptoms of others (Chapter five). To date, most studies of the effects of within-host microbial interactions are at the individual level, focusing on the responses of parasites to coinfecting microbes within an individual host. Studies

of population-level responses to microbial interactions are currently limited; conducting more such studies would increase understanding of how microbial interactions affect evolutionary responses, which in turn will give new insight into how virulent parasites arise. Furthermore, integrating individual-level and population-level studies of the same system, as this dissertation did, would provide much needed insight on the connection between the ecology and evolution of infectious diseases. Such integration would be invaluable when considering short-term and long-term results of disease management efforts related to emerging diseases and pathogens expanding their host range. Foliar fungal parasites offer a tractable system to apply such a paradigm. With more population-level studies looking at the evolutionary effects of within-host microbial interactions, and the integration of this population-level data with data at the individual level, we will be able to improve ecological and evolutionary understanding of parasites.

APPENDIX A: SUPPLEMENTARY MATERIAL FOR CHAPTER 2

A1. Supplemental Figures

Figure A2.1: Trinity assembly metrics. The blue points in the top graph indicate the number of contigs in the assembly of each simulated dataset. Assemblies are comprised of fewer contigs when datasets are closer to being composed of just one species. The red points of the bottom graph indicate the mean length of the contigs of each assembly.

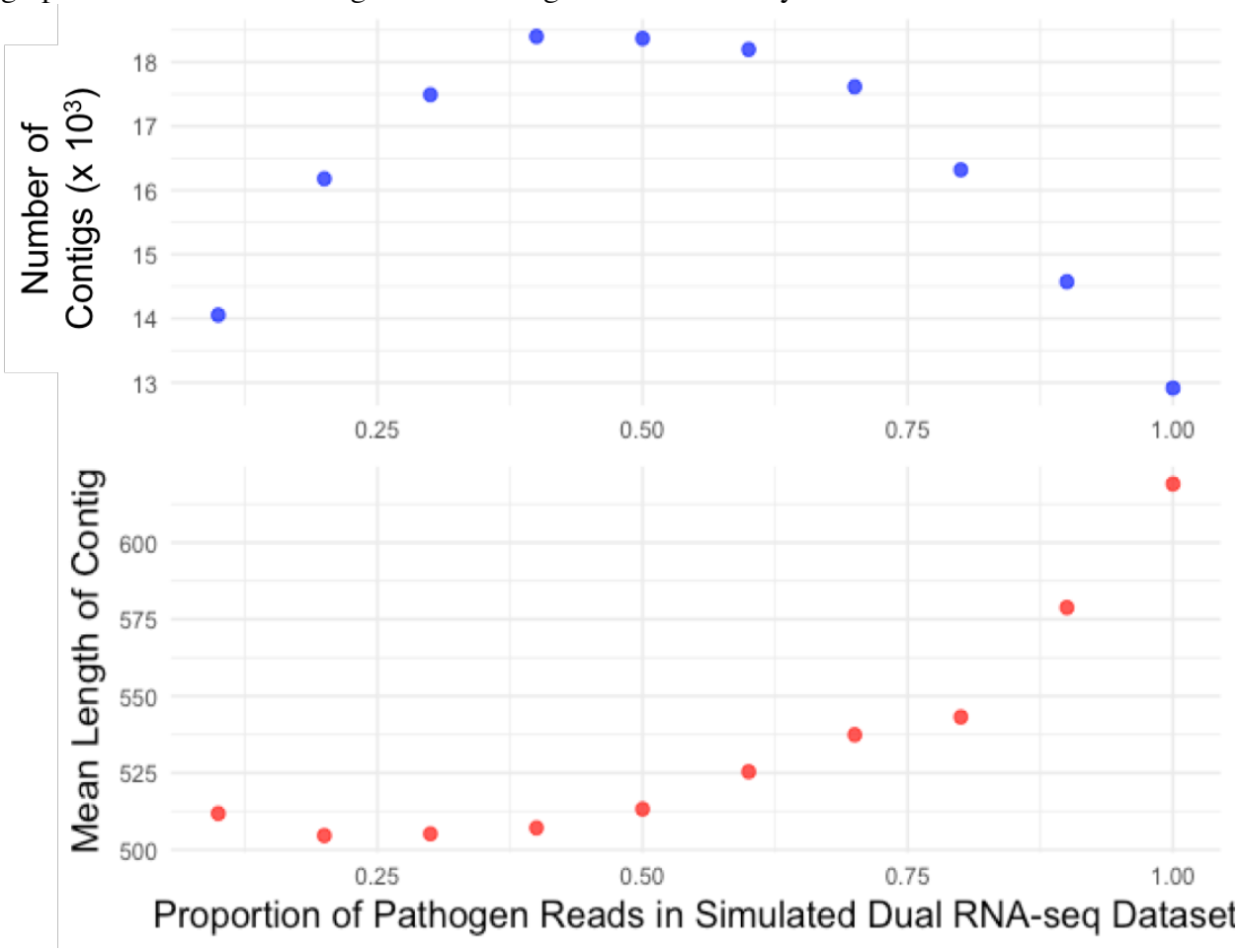


Figure A2.2: K-mer counts of Sequencing Reads Included (A) or Excluded (B) from Trinity Assembles of 76 single-end datasets. Approximately 8.7 reads were included in each assembly and approximately 1.3 million reads were excluded from the assemblies of each dataset. Most of the unused reads are rare, but a small subset are highly repetitive.

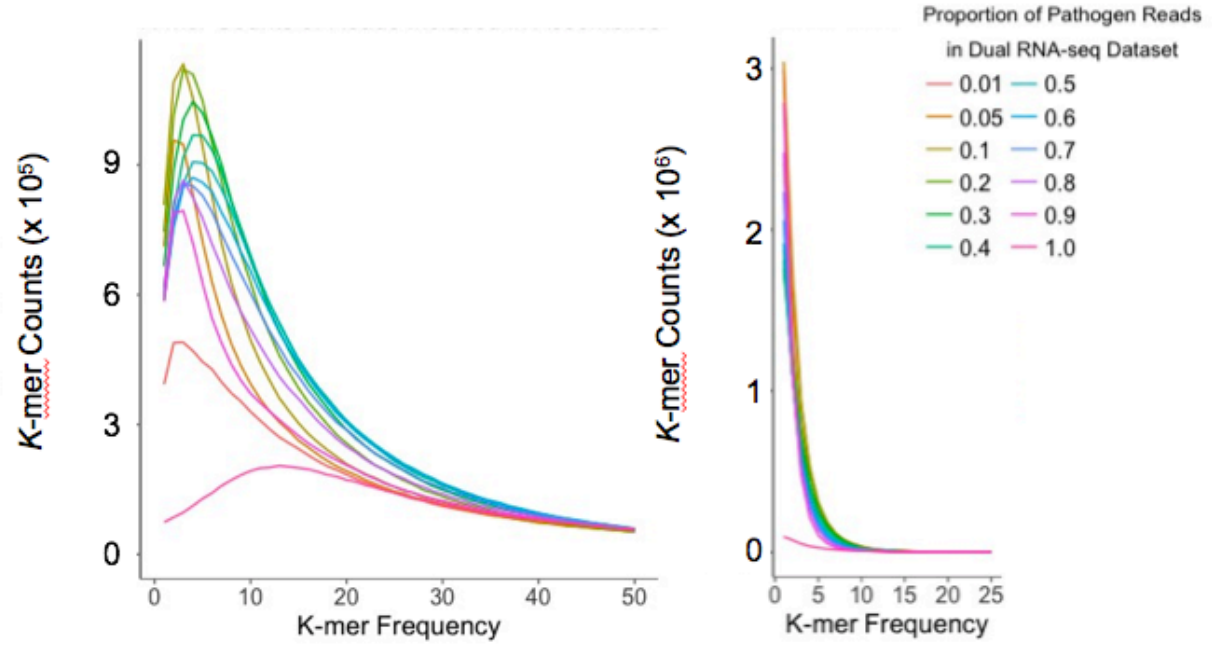


Figure A2.3: Comparison of the utility and accuracy of several aligners for dual RNA-seq analysis of datasets with very low proportions of pathogen reads. Results are shown for a factorial combination of four aligners and the four *Schizosaccharomyces* reference genomes. For each bar plot, blue reads are those that originated from the host (*Arabidopsis thaliana*) and yellow reads are those that originated from the pathogen (*Schizosaccharomyces octosporus*). TopHat2 and MapSplice2 can only effectively map pathogen reads when mapping to the reference genome of the pathogen, and are unable to effectively map reads to the reference genomes of related species. These plots are therefore excluded. STAR and NextGenMap both result in host read mismapping when there are very low proportions of pathogen reads in datasets.

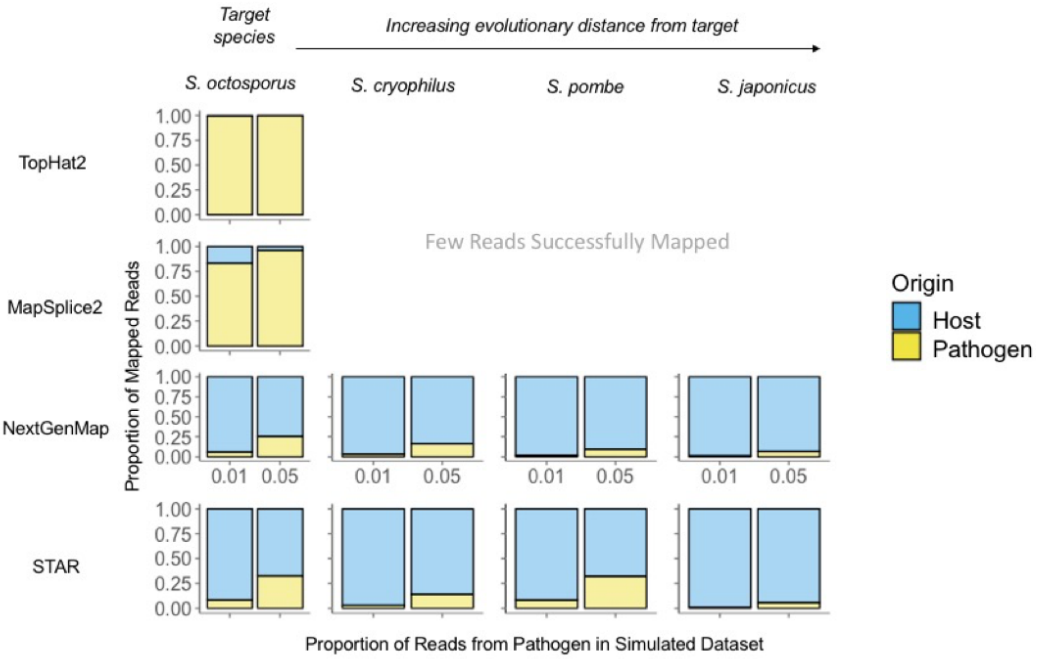


Figure A2.4 Concatenated genome mapping improve dual RNA-seq analysis with NextGenMap of datasets very low proportions of pathogen reads. Results are shown for aligners STAR and NextGenMap and the four composite genomes comprised of the *Arabidopsis thaliana* (host) genome and each of four *Schizosaccharomyces* reference genomes. For each bar plot, blue reads are those that originated from the host (*Arabidopsis thaliana*) and yellow reads are those that originated from the target pathogen (*Schizosaccharomyces octosporus*). The bar plots represent the composition of the reads that mapped to the component of each composite genome corresponding to either the target pathogen genome or the genome of a closely-related species. Host mismapping was strongly reduced by mapping with NextGenMap to a concatenated genome, while for datasets with very low proportions of pathogen reads, STAR still experiences host read mismapping.

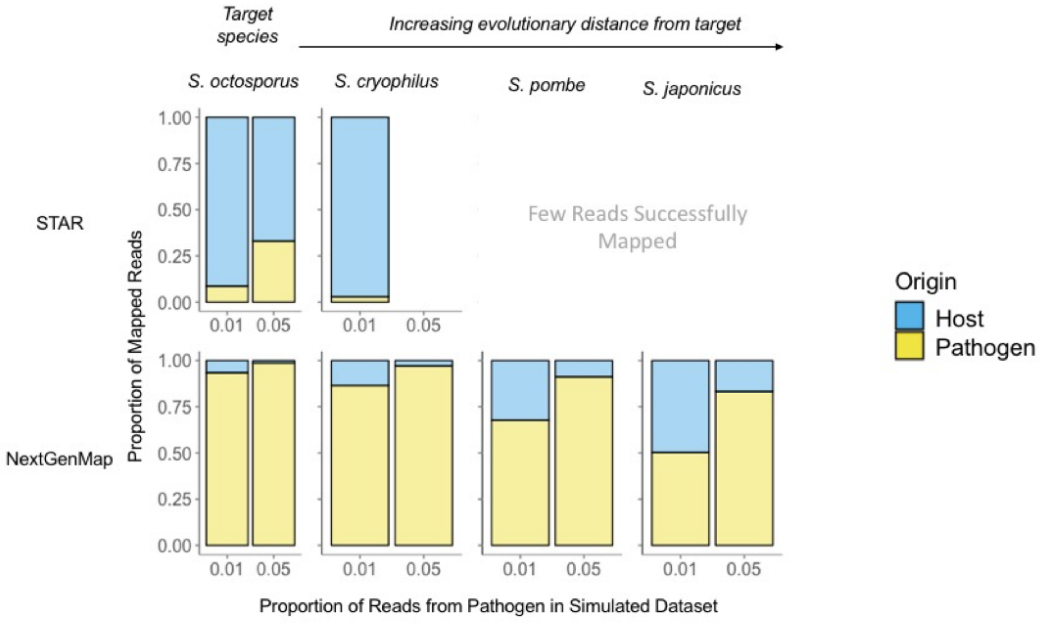


Figure A2.5 Assembling reads *de novo* first improves alignment of dual RNA-seq datasets very low proportions of pathogen reads. Results are shown for aligners STAR and NextGenMap and the four *Schizosaccharomyces* reference genomes. For each bar plot, blue contigs are those that originated from the host (*Arabidopsis thaliana*) and yellow contigs are those that originated from the pathogen (*Schizosaccharomyces octosporus*). The bar plots represent the composition of the contigs that mapped to each reference genome. STAR was unable to align more than a few contigs for *S. cryophilus*, *S.pombe*, and *S. japonicus*, so those plots have been excluded. Host mismapping was strongly reduced.

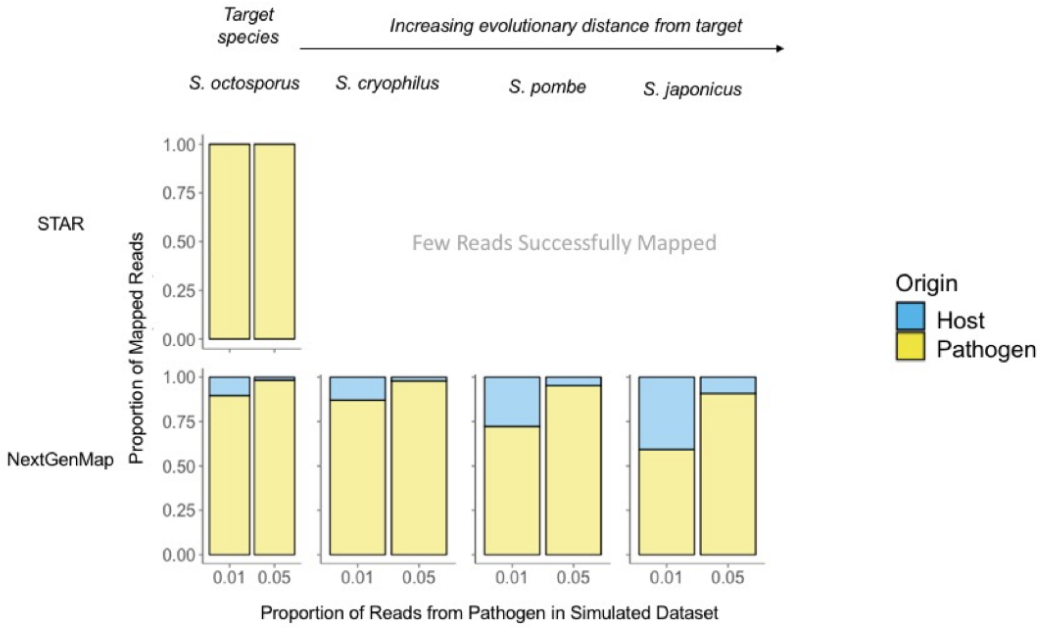


Figure A2.6 Comparison of the utility and accuracy of several aligners for dual RNA-seq analysis in human-*Candida* system. Results are shown for a factorial combination of 3 aligners and the three *Candida* reference genomes. A. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. TopHat2 can only map pathogen reads when mapping to the reference genome of the pathogen and is unable to effectively map reads to the reference genomes of related species. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species. B. For each bar plot, blue contigs are those that originated from the host and yellow contigs are those that originated from the pathogen. Because TopHat2 is unable to map a considerable number of reads to the genomes of species related to the pathogen, these plots are excluded. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species, but both aligners result in host reads mismapping.

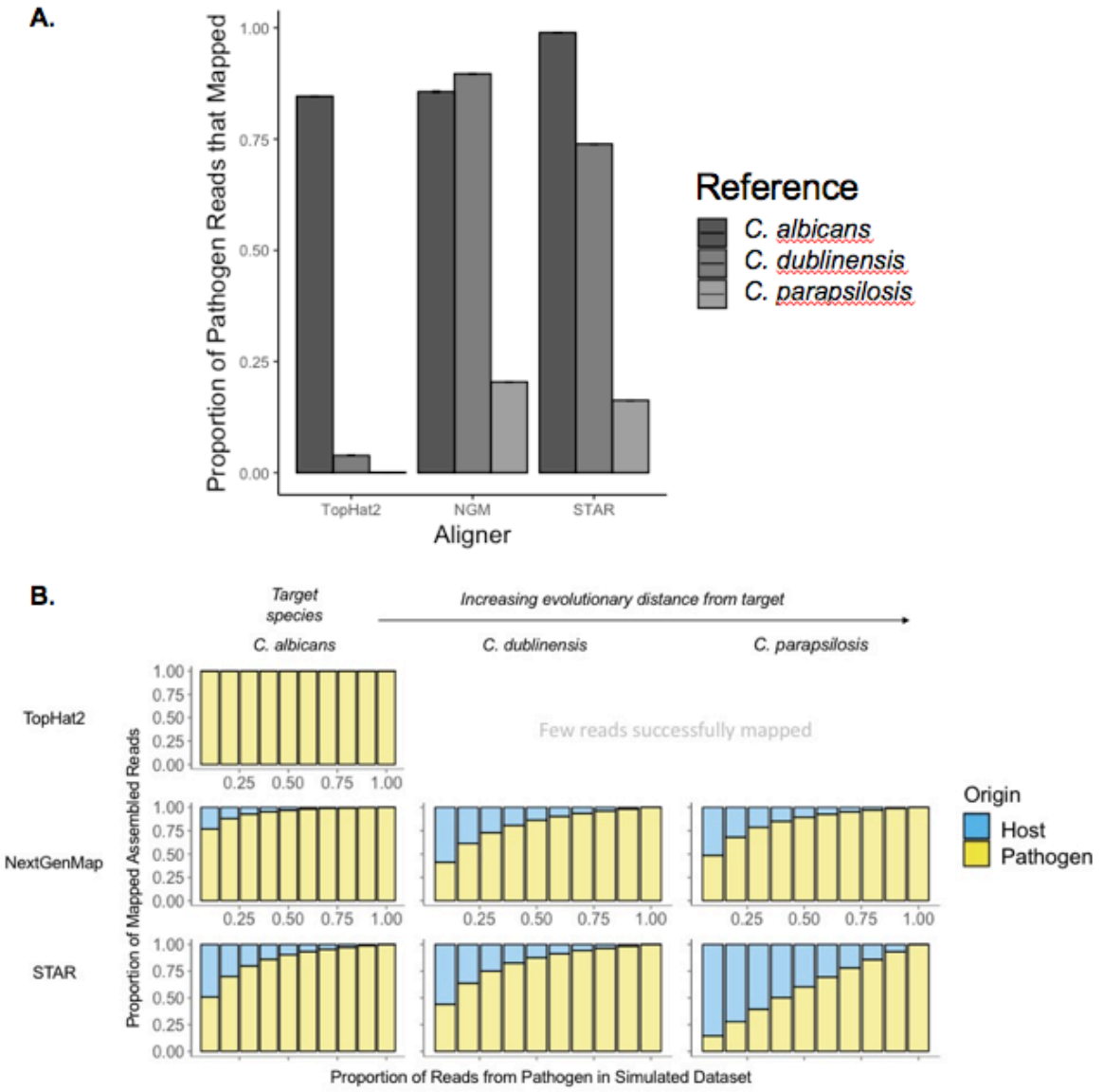


Figure A2.7 Mapping Reads to Host Genome to Filter Host Reads Does Not Reduce Host Read Mismapping (Human-*Candida*)

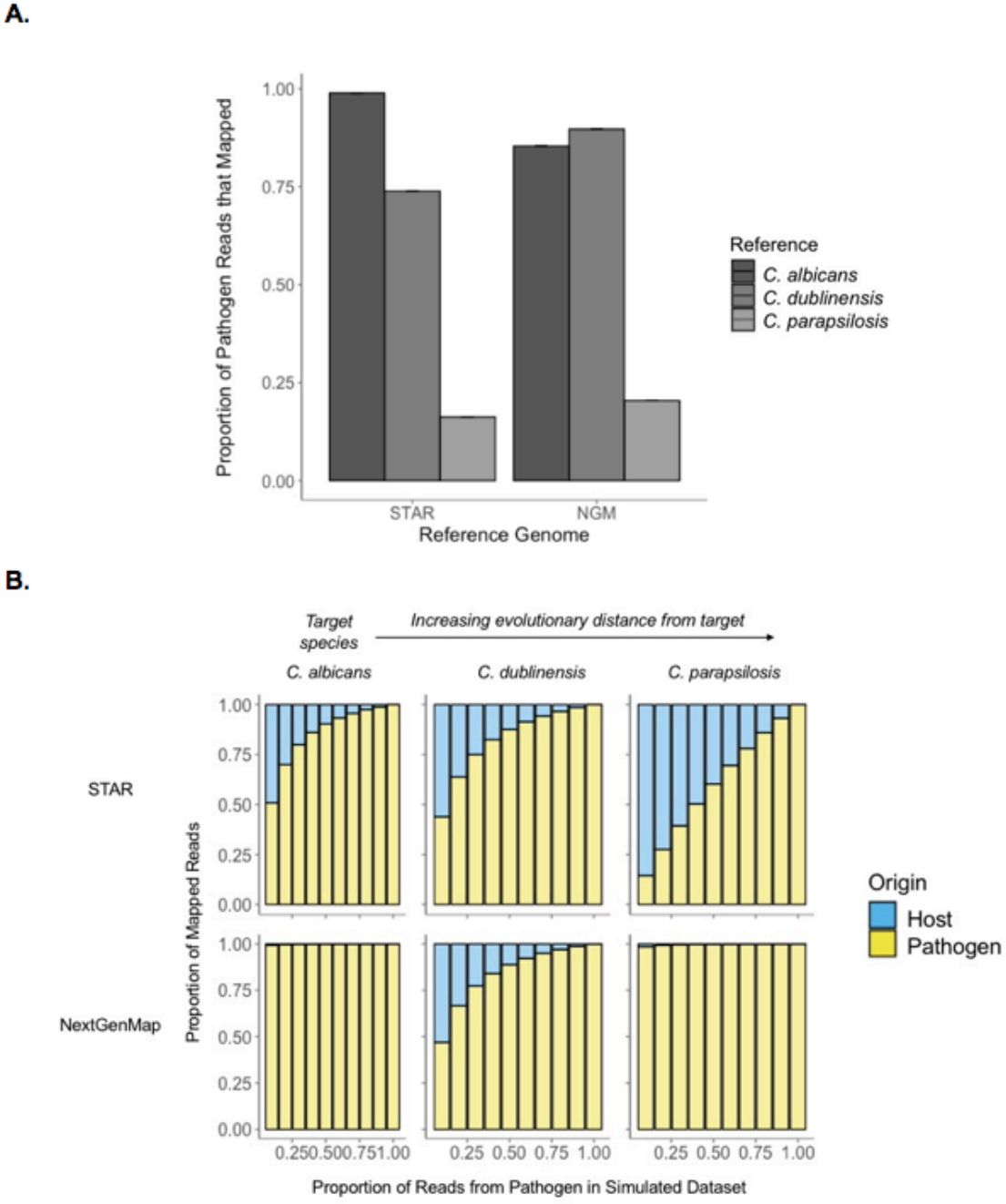


Figure A2.8 Assembling reads *de novo* first improves alignment of dual RNA-seq datasets of Human-*Candida albicans*. Results are shown for aligners STAR and NextGenMap and the three *Candida* reference genomes. A. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species, though NextGenMap has a much higher mapping rate of pathogen reads overall B. For each bar plot, blue contigs are those that originated from the host, yellow contigs are those that originated from the pathogen, and green contigs are those that could not be identified as from the host or pathogen. The bar plots represent the composition of the contigs that mapped to each reference genome. STAR was unable to align more than a few contigs for *C. parapsilosis,* so that plot has been excluded. Host mismapping was strongly reduced.

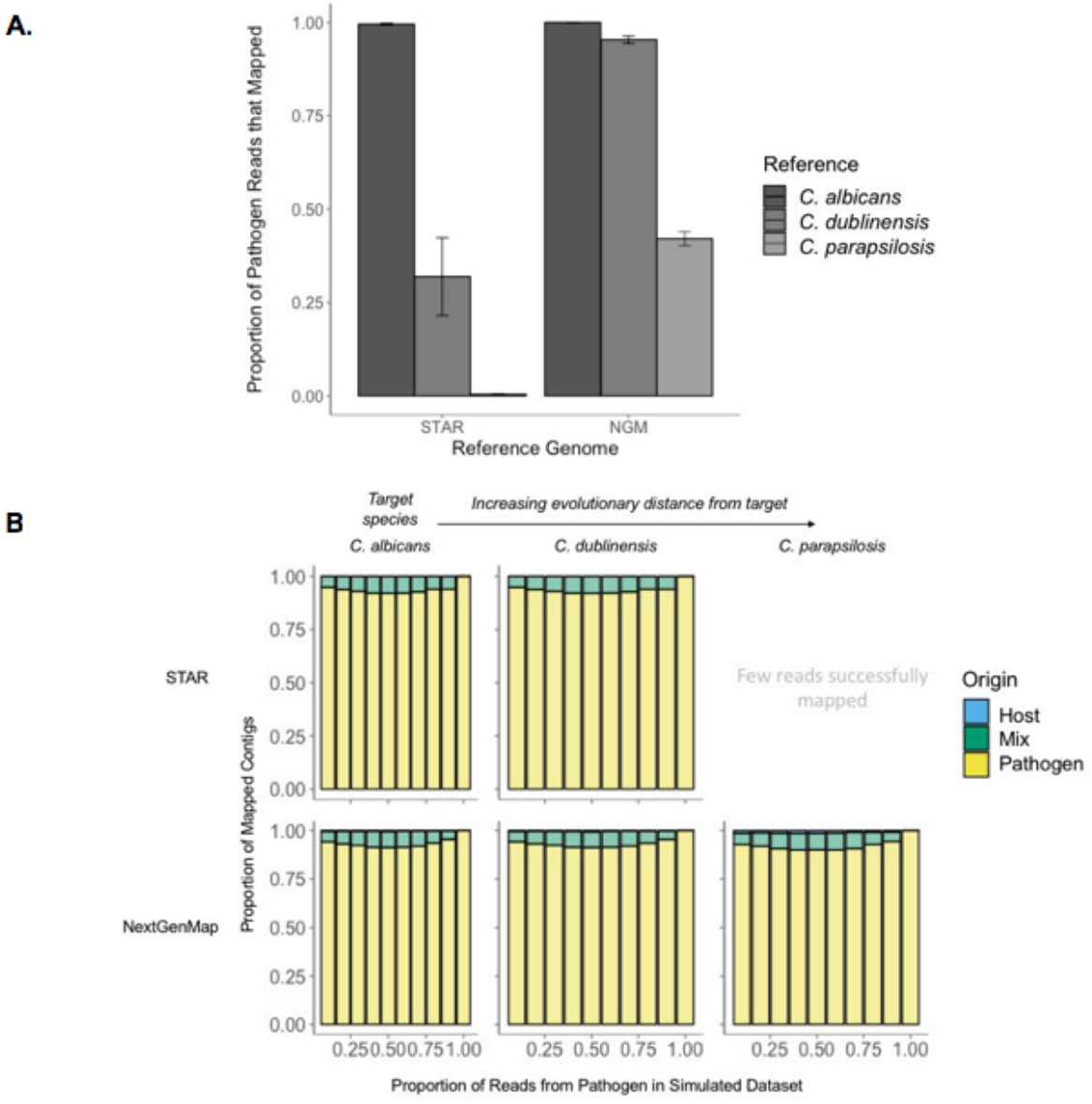


Figure A2.9 Concatenated genome mapping improve dual RNA-seq analysis of Human-*Candida* datasets. Results are shown for aligners STAR and NextGenMap and the three concatenated genomes comprised of the host genome (*Homo sapiens)* and each of three *Candida* reference genomes. A. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species, though NextGenMap has a much higher mapping rate of pathogen reads overall B. For each bar plot, blue reads are those that originated from the host (*Homo sapiens*) and yellow reads are those that originated from the target pathogen (*Candida albicans*). The bar plots represent the composition of the reads that mapped to the component of each concatenated genome corresponding to either the target pathogen genome or the genome of a closely-related species. Host mismapping was strongly reduced.

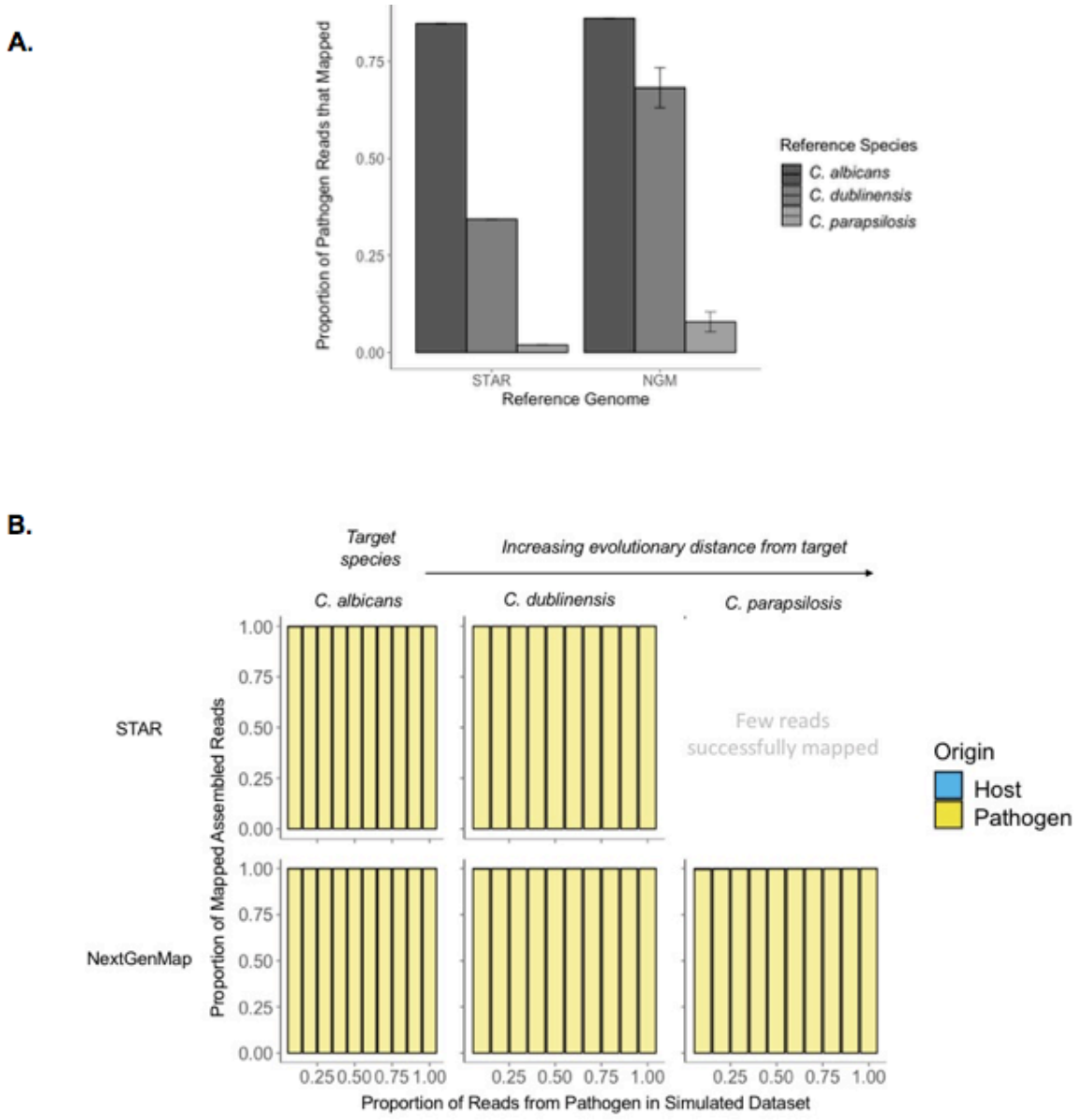


Figure A2.10 Comparison of the utility and accuracy of several aligners for dual RNA-seq analysis in human-*E. coli* system. Results are shown for a factorial combination of 3 aligners and the three *Escherichia* reference genomes. A. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. TopHat2 can only map pathogen reads when mapping to the reference genome of the pathogen and is unable to effectively map reads to the reference genomes of related species. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species. B. For each bar plot, blue contigs are those that originated from the host and yellow contigs are those that originated from the pathogen. Because TopHat2 is unable to map a considerable number of reads to the genomes of species related to the pathogen, these plots are excluded. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species, but both aligners result in host reads mismapping.

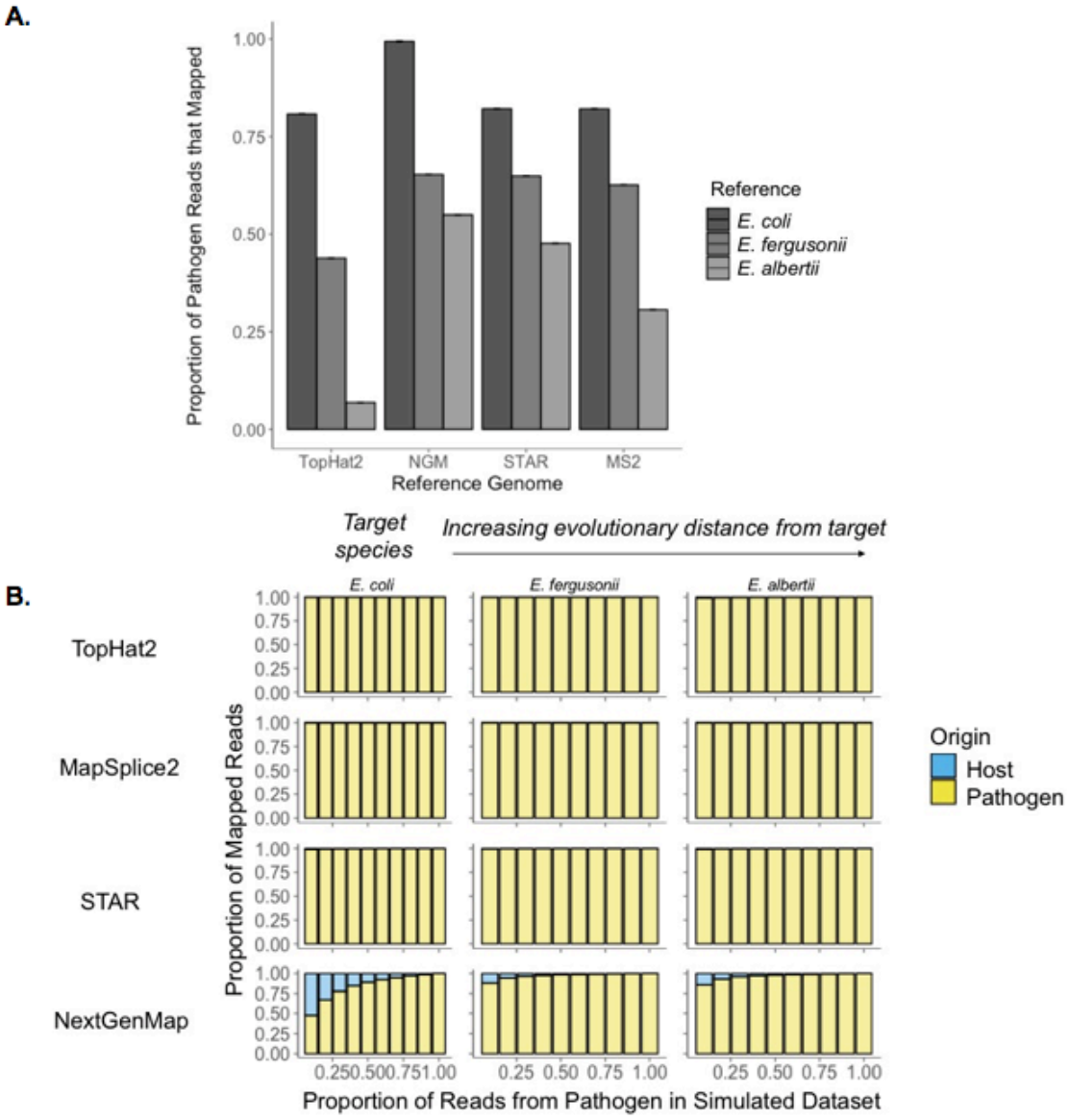


Figure A2.11 Comparison of the utility and accuracy of several aligners for dual RNA-seq (150 single-end datasets) analysis. Results are shown for a factorial combination of four aligners and the four *Schizosaccharomyces* reference genomes. A. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. TopHat2 and STAR can only map pathogen reads when mapping to the reference genome of the pathogen, and is unable to effectively map reads to the reference genomes of related species. NextGenMap is able to map pathogen reads to the reference genomes of related species. B. For each bar plot, blue reads are those that originated from the host (*Arabidopsis thaliana*) and yellow reads are those that originated from the pathogen (*Schizosaccharomyces octosporus*). TopHat2 and STAR can only effectively map pathogen reads when mapping to the reference genome of the pathogen and in the case of STAR, that of the most closely-related species, and are unable to effectively map reads to the other reference genomes. These plots are therefore excluded. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species, but both aligners result in host reads mismapping.

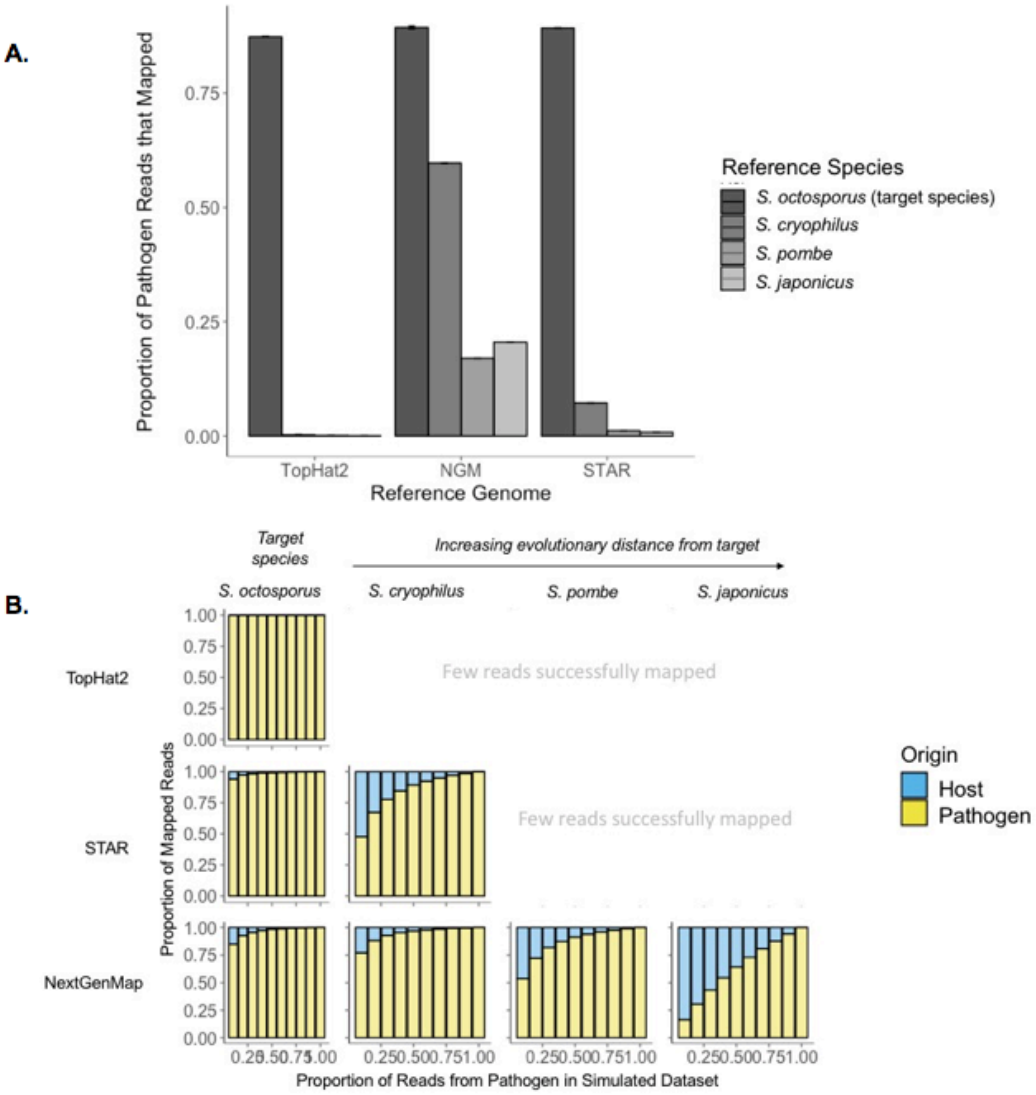


Figure A2.12 Comparison of the utility and accuracy of several aligners for dual RNA-seq (76 paired-datasets) analysis. Results are shown for a factorial combination of four aligners and the four *Schizosaccharomyces* reference genomes. A. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. TopHat2 and STAR can only map pathogen reads when mapping to the reference genome of the pathogen, and is unable to effectively map reads to the reference genomes of related species. NextGenMap is able to map pathogen reads to the reference genomes of related species. B. For each bar plot, blue reads are those that originated from the host (*Arabidopsis thaliana*) and yellow reads are those that originated from the pathogen (*Schizosaccharomyces octosporus*). TopHat2 and STAR can only effectively map pathogen reads when mapping to the reference genome of the pathogen and in the case of STAR, that of the most closely-related species, and are unable to effectively map reads to the other reference genomes. These plots are therefore excluded. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species, but both aligners result in host reads mismapping.

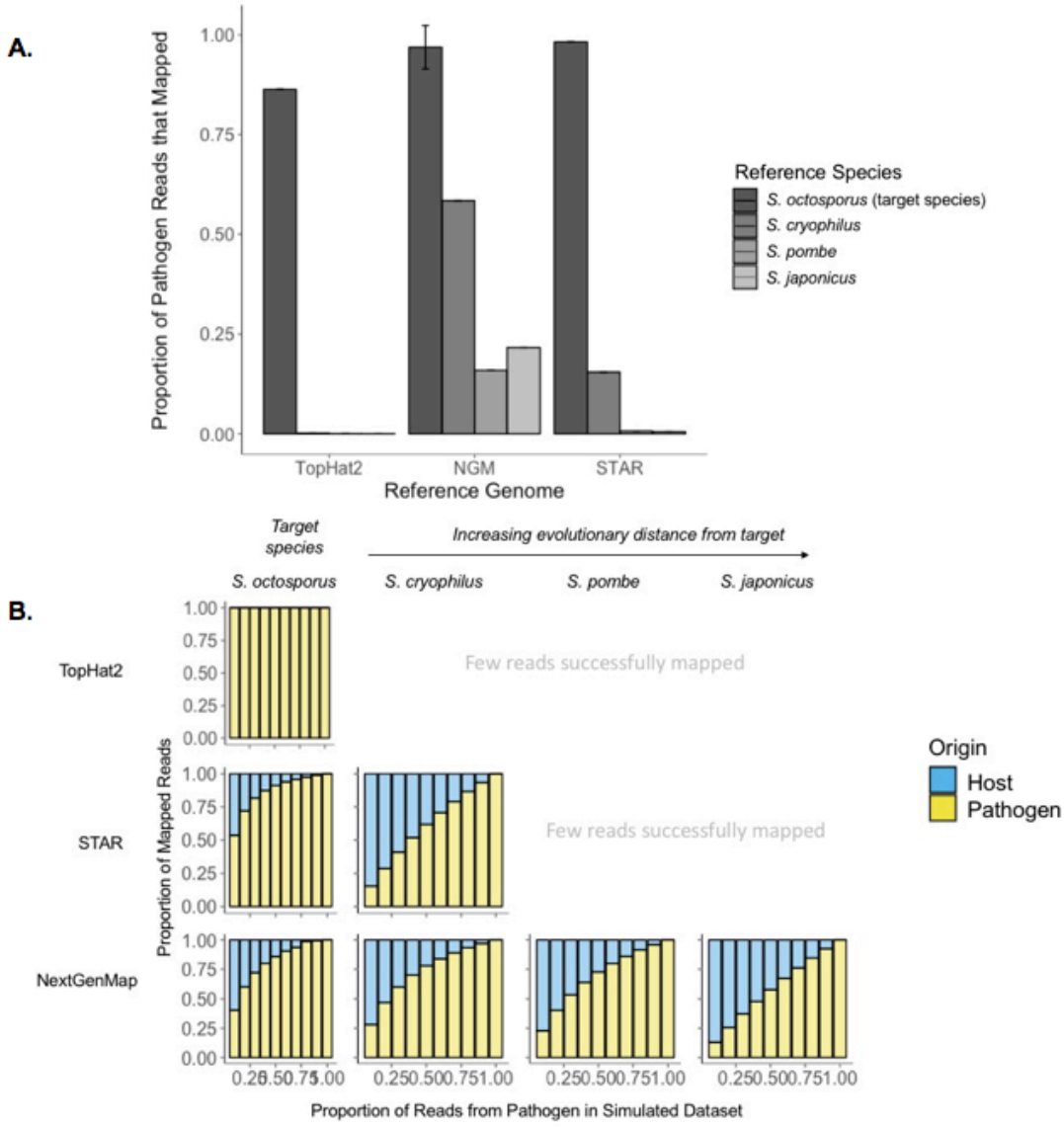


Figure A2.13 Assembling reads *de novo* first improves alignment of dual RNA-seq 150 single-end reads. Results are shown for aligners STAR and NextGenMap and the four *Schizosaccharomyces* reference genomes. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. STAR can only map pathogen contigs when mapping to the reference genome of the pathogen and is unable to effectively map contigs to the reference genomes of related species. NextGenMap retains its ability to map pathogen contigs to transcriptomes of related species. B. For each bar plot, blue contigs are those that originated from the host (*Arabidopsis thaliana*) and yellow contigs are those that originated from the pathogen (*Schizosaccharomyces octosporus*). Contigs that were unable to be determined as comprised of host or pathogen reads are colored green. The bar plots represent the composition of the contigs that mapped to each reference genome. STAR was unable to align more than a few contigs for *S. cryophilus*, *S.pombe*, and *S. japonicus*, so those plots have been excluded. Host mismapping was strongly reduced.

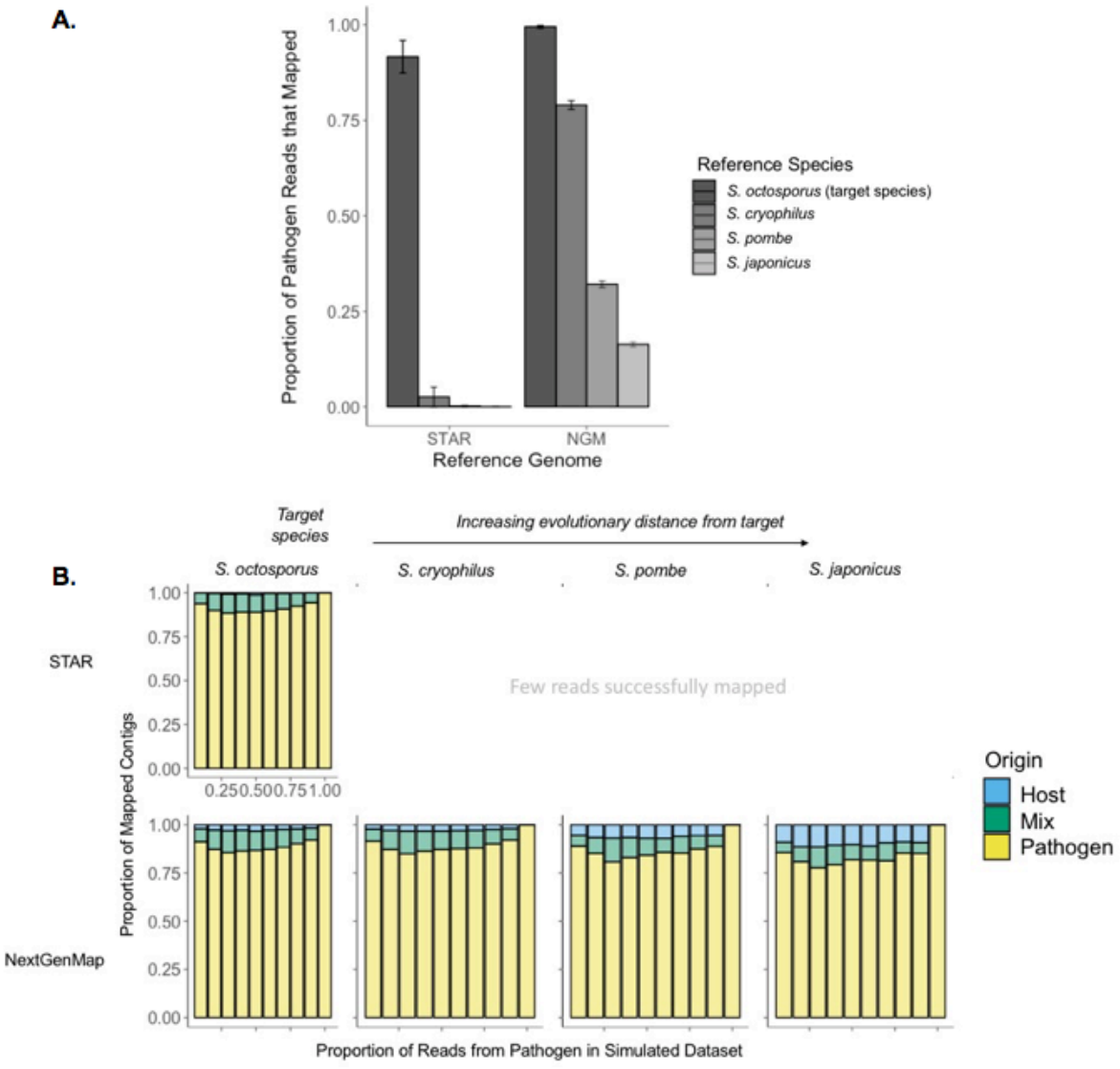


Figure A2.14 Assembling reads *de novo* first improves alignment of dual RNA-seq 76 paired-end reads. Results are shown for aligners STAR and NextGenMap and the four *Schizosaccharomyces* reference genomes. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. STAR can only map pathogen contigs when mapping to the reference genome of the pathogen and is unable to effectively map contigs to the reference genomes of related species. NextGenMap retains its ability to map pathogen contigs to transcriptomes of related species. B. For each bar plot, blue contigs are those that originated from the host (*Arabidopsis thaliana*) and yellow contigs are those that originated from the pathogen (*Schizosaccharomyces octosporus*). Contigs that were unable to be determined as comprised of host or pathogen reads are colored green. The bar plots represent the composition of the contigs that mapped to each reference genome. STAR was unable to align more than a few contigs for *S. cryophilus*, *S.pombe*, and *S. japonicus*, so those plots have been excluded. Host mismapping was strongly reduced.

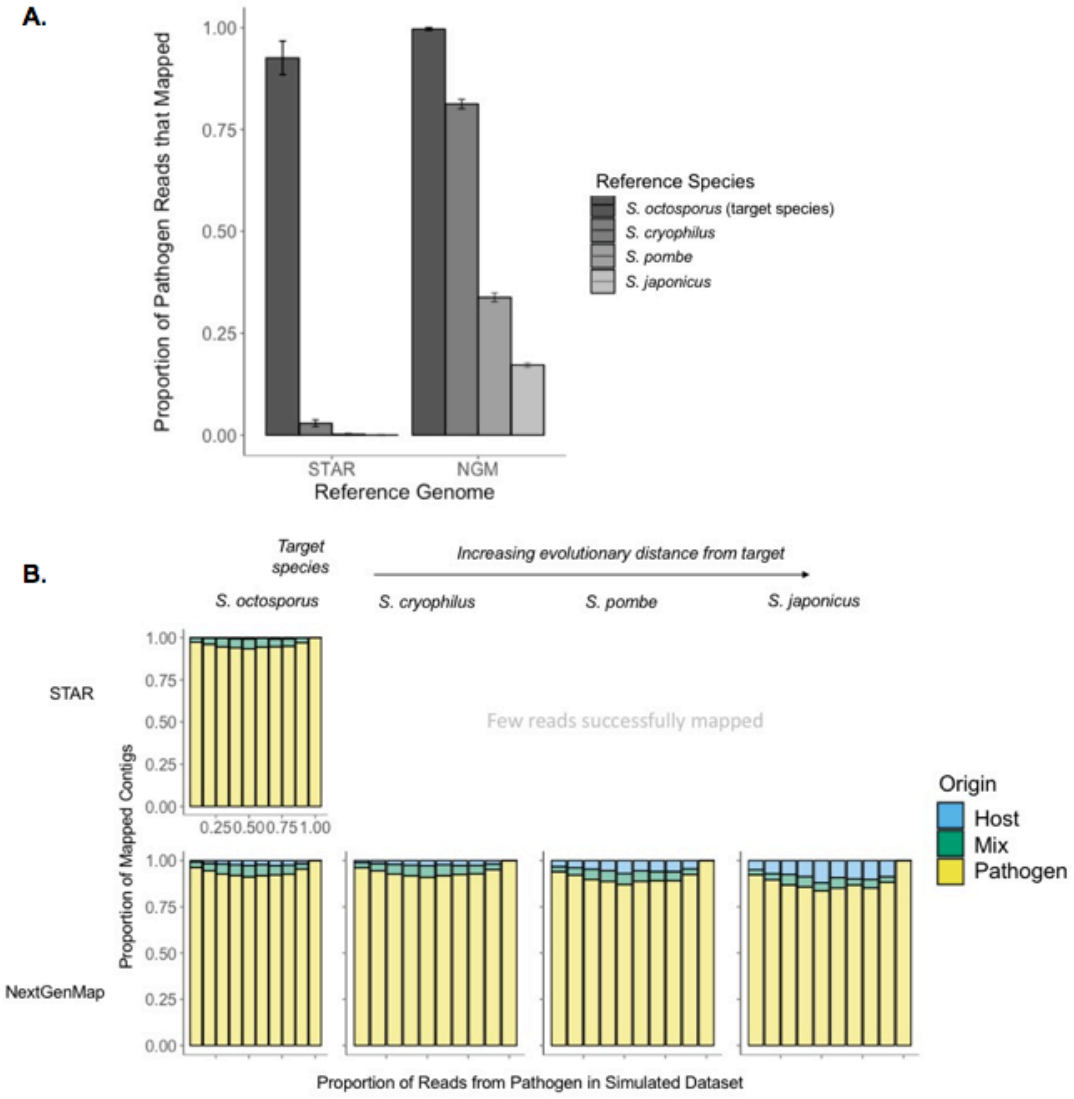


Figure A2.15 Concatenated genome mapping improves dual RNA-seq analysis of 150 single-end datasets. Results are shown for aligners STAR and NextGenMap and the four concatenated genomes comprised of the host genome (*Arabidopsis thaliana)* and each of four *Schizosaccharomyces* reference genomes. A. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species, though NextGenMap has a much higher mapping rate of pathogen reads overall B. For each bar plot, blue reads are those that originated from the host (*A. thaliana*) and yellow reads are those that originated from the target pathogen (*S. octosporus*). The bar plots represent the composition of the reads that mapped to the component of each concatenated genome corresponding to either the target pathogen genome or the genome of a closely-related species. Host mismapping was strongly reduced, especially with NextGenMap.

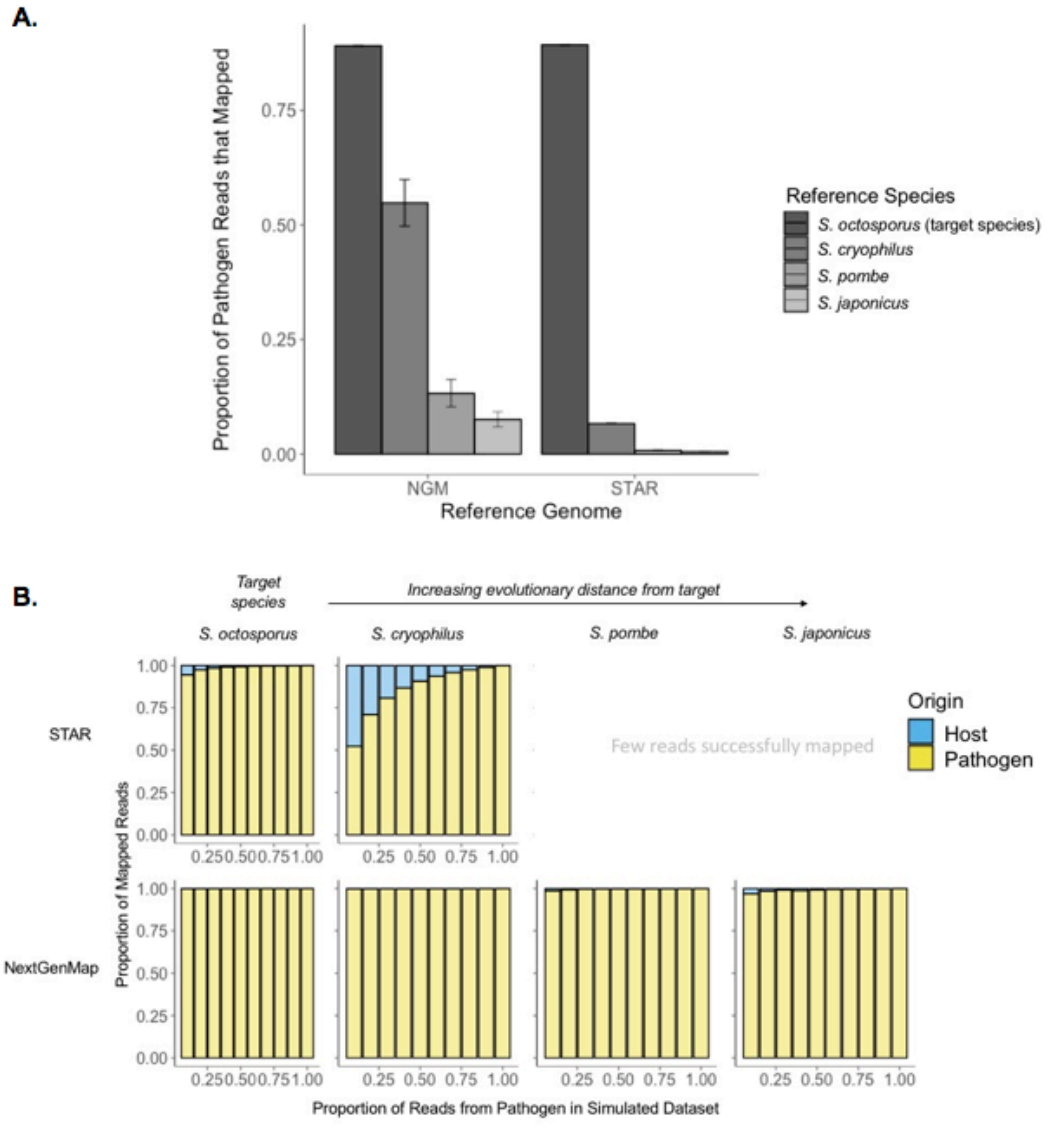
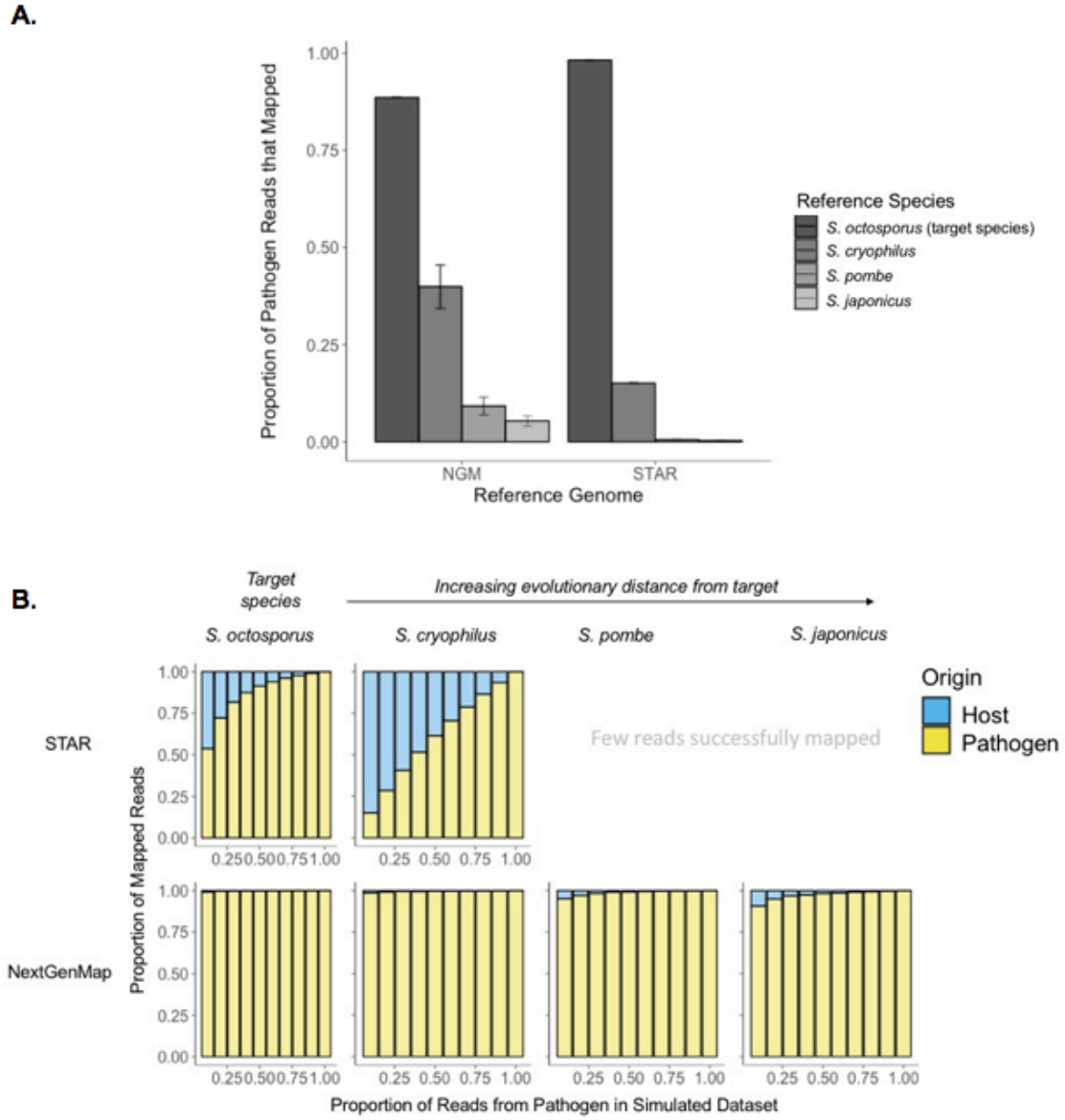


Figure A2.16 Concatenated genome mapping improves dual RNA-seq analysis of 76 paired-end datasets. Results are shown for aligners STAR and NextGenMap and the four concatenated genomes comprised of the host genome (*Arabidopsis thaliana)* and each of four *Schizosaccharomyces* reference genomes. A. Bars indicate the proportion of pathogen reads that mapped to each genome/transcriptome. Alignments to the target pathogen species are shown in black, and the grayscale gradient indicates evolutionary distance from the target. STAR and NextGenMap are able to map pathogen reads to the reference genomes of related species, though NextGenMap has a much higher mapping rate of pathogen reads overall. B. For each bar plot, blue reads are those that originated from the host (*A. thaliana*) and yellow reads are those that originated from the target pathogen (*S. octosporus*). The bar plots represent the composition of the reads that mapped to the component of each concatenated genome corresponding to either the target pathogen genome or the genome of a closely-related species. Host mismapping was strongly reduced with NextGenMap.



A2. Supplemental Tables

Table A2.1: Schizosaccharomyces Species Information

Species	**1:1 Ortholog Amino Acid Identity to Target***	**Genome Size**	**GC Content**
*S. octosporus**	-	11.5 Mb	38%
S. cryophilus	85%	12.5 Mb	38%
S. pombe	66%	12.5 Mb	36%
S. japonicus	56%	12.5 Mb	44%

*Target Species: Species used for dual RNA-seq simulations

Table A2.2: Species Information for Pathogens Included in simulations, as well as closely-related species used for reference genomes

Species	Genome Size	Number of Genes	GC Content
*C. albicans**	14.4 Mb	6159	33.50%
C. dubliniensis	14.6 Mb	5758	33.10%
C. parapsilosis	13.1Mb	5733	38.70%

Species	Genome Size	Number of Genes	GC Content
*E. coli**	4.6 Mb	4288	50.60%
E. fergusonii	4.6 Mb	4641	49.90%
E. albertii	4.7 Mb	4533	49.90%

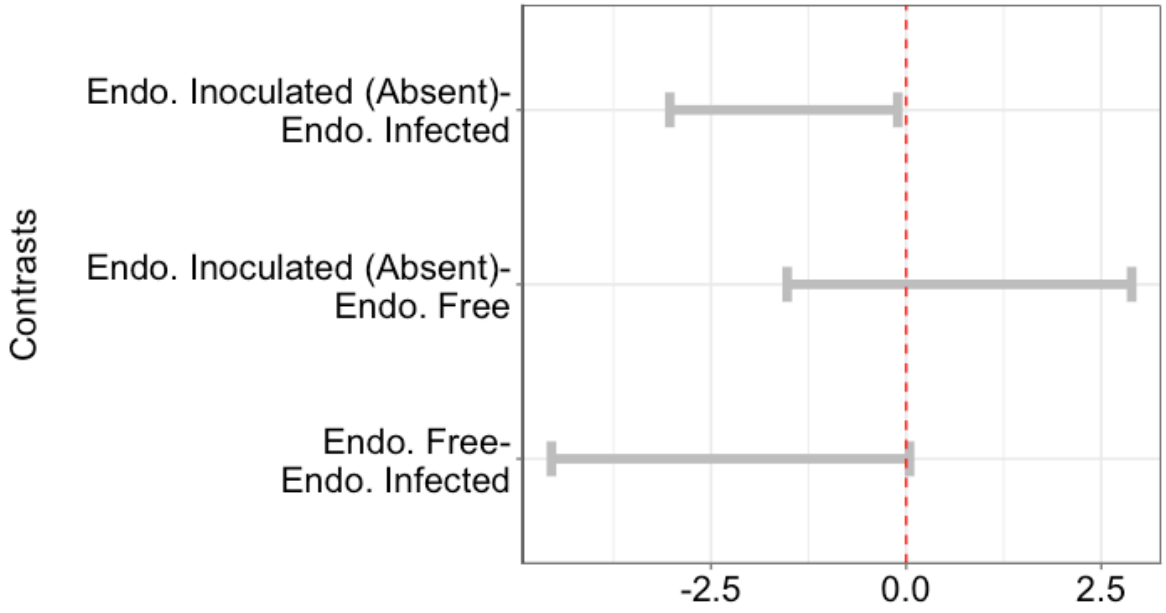
*Target Species: Species used for dual RNA-seq simulations

Table A2.3: Extent of Host Read Mismapping

	Number of Genes	Percent of total genes
Genes Where 1 or more host read maps	4246	36.96674212
Genes Where 5 or more host read maps	1898	16.52446457
Genes Where 20 or more host read maps	558	4.858088107
Genes Where 50 or more host read maps	192	1.671600209
Total genes	11486	

APPENDIX B: SUPPLEMENTARY MATERIAL FOR CHAPTER 4

Figure B4.1 Tukey Simultaneous 95% Confidence Intervals for differences of means for endophyte infection category



APPENDIX C: SUPPLEMENTARY MATERIAL FOR CHAPTER 5

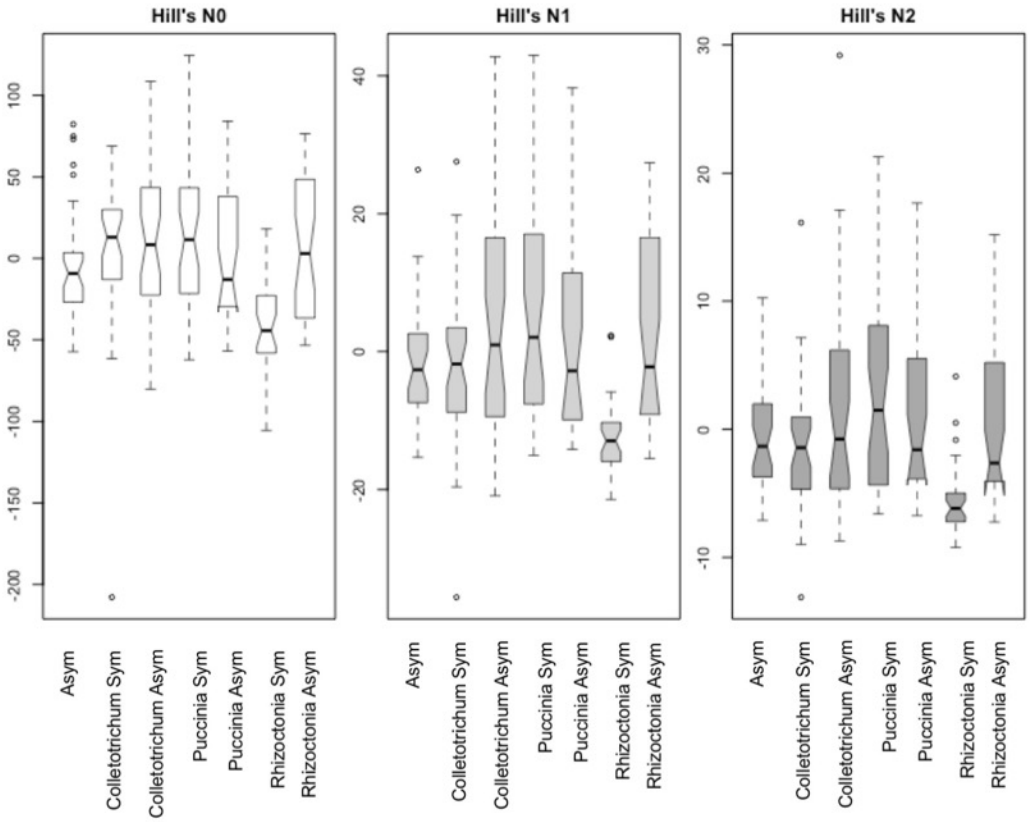


Figure C5.1 Leaf segments with symptoms of *Rhizoctonia solani* were associated with less diverse foliar fungal communities. Panels show partial residuals of Hill's diversity series after regressing out differential sequencing depth in linear mixed models. Leaf segments with *R. solani* lesions had statistically significantly lower diversity in comparison to the other treatments, evaluated with Tukey's HSD.

REFERENCES

Abbate, J. L., Ezenwa, V. O., Guégan, J. F., Choisy, M., Nacher, M., & Roche, B. (2018). Disentangling complex parasite interactions: Protection against cerebral malaria by one helminth species is jeopardized by co-infection with another. *PLoS Neglected Tropical Diseases*, *12*(5), 1–13. doi:10.1371/journal.pntd.0006483

Adame-Álvarez, R.-M., Mendiola-Soto, J., & Heil, M. (2014a). Order of arrival shifts endophyte-pathogen interactions in bean from resistance induction to disease facilitation. *FEMS Microbiology Letters*, *355*(2), 100–7. doi:10.1111/1574-6968.12454

Adame-Álvarez, R.-M., Mendiola-Soto, J., & Heil, M. (2014b). Order of arrival shifts endophyte–pathogen interactions in bean from resistance induction to disease facilitation. *FEMS Microbiology Letters*, *355*(2), 100–107. doi:10.1111/1574-6968.12454

Aivelo, T., & Norberg, A. (2018). Parasite–microbiota interactions potentially affect intestinal communities in wild mammals. *Journal of Animal Ecology*, *87*(2), 438–447. doi:10.1111/1365-2656.12708

Al-Naimi, F. a, Garrett, K. a, & Bockus, W. W. (2005). Competition, facilitation, and niche differentiation in two foliar pathogens. *Oecologia*, *143*(3), 449–57. doi:10.1007/s00442-004-1814-x

Alizon, S., de Roode, J. C., & Michalakis, Y. (2013). Multiple infections and the evolution of virulence. *Ecology Letters*, *16*(4), 556–67. doi:10.1111/ele.12076

Aprianto, R., Slager, J., Holsappel, S., & Veening, J.-W. (2016). Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. *Genome Biology*, *17*(1), 198. doi:10.1186/s13059-016-1054-5

Arnold, a. E., Henk, D. a., Eells, R. L., Lutzoni, F., & Vilgalys, R. (2007). Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR. *Mycologia*, *99*(2), 185–206. doi:10.3852/mycologia.99.2.185

Arnold, A. E., Mejía, L. C., Kyllo, D., Rojas, E. I., Maynard, Z., Robbins, N., & Herre, E. A. (2003). Fungal endophytes limit pathogen damage in a tropical tree. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15649–15654. doi:10.1073/pnas.2533483100

Baddal, B., Muzzi, A., Censini, S., Calogero, R., Torricelli, G., Guidotti, S., … Paxxicoli, A. (2015). Dual RNA-seq of Nontypeable Haemophilus influenzae and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk. *MBio*, *6*(e01765-15). doi:10.1128/mBio.01765-15

Barman, M., Unold, D., Shifley, K., Amir, E., Hung, K., Bos, N., & Salzman, N. (2008). Enteric salmonellosis disrupts the microbial ecology of the murine gastrointestinal tract. *Infection*

and Immunity, *76*(3), 907–915. doi:10.1128/IAI.01432-07

Baruzzo, G., Hayer, K. E., Kim, E. J., Di Camillo, B., FitzGerald, G. A., & Grant, G. R. (2016). Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nature Methods*, *14*, 135.

Baxter, N. T., Wan, J. J., Schubert, A. M., Jenior, M. L., Myers, P., & Schloss, P. D. (2015). Intra- and interindividual variations mask interspecies variation in the microbiota of sympatric Peromyscus populations. *Applied and Environmental Microbiology*, *81*(1), 396–404. doi:10.1128/AEM.02303-14

Beirn, L. A., Wang, R., Clarke, B. B., & Crouch, J. A. (2015). Development of a greenhouse-based inoculation protocol for the fungus Colletotrichum cereale pathogenic to annual bluegrass (Poa annua). *PeerJ*, *3*, e1153–e1153. doi:10.7717/peerj.1153

Belkaid, Y., & Hand, T. (2014). Role of microbiota in immunity and inflammation. *Cell*, *157*(1), 121–141. doi:10.1016/j.cell.2014.03.011.

Benjamin, A. M., Nichols, M., Burke, T. W., Ginsburg, G. S., & Lucas, J. E. (2014). Comparing reference-based RNA-Seq mapping methods for non-human primate data. *BMC Genomics*, *15*(1), 570. doi:10.1186/1471-2164-15-570

Bent, S. J., & Forney, L. J. (2008). The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *The Isme Journal*, *2*, 689.

Berg, M., & Koskella, B. (2018). Nutrient- and Dose-Dependent Microbiome-Mediated Protection against a Plant Pathogen. *Current Biology*, *28*(15), 2487–2492.e3. doi:https://doi.org/10.1016/j.cub.2018.05.085

Bonos, S. A., Wilson, M. M., Meyer, W. A., & Funk, C. R. (2005). Suppression of Red Thread in Fine Fescues Through Endophyte-Mediated Resistance. *Applied Turfgrass Science*, *2*.

Bordes, F., & Morand, S. (2011). The impact of multiple infections on wild animal hosts: a review. *Infection Ecology & Epidemiology*, *1*(1), 7346. doi:10.3402/iee.v1i0.7346

Borer, E. T., Kinkel, L. L., May, G., & Seabloom, E. W. (2013a). The world within: Quantifying the determinants and outcomes of a host's microbiome. *Basic and Applied Ecology*, *14*(7), 533–539. doi:10.1016/j.baae.2013.08.009

Borer, E. T., Kinkel, L. L., May, G., & Seabloom, E. W. (2013b). The world within: Quantifying the determinants and outcomes of a host's microbiome. *Basic and Applied Ecology*, *14*(7), 533–539. doi:http://doi.org/10.1016/j.baae.2013.08.009

Bronstein, J. L. (2001). The exploitation of mutualisms. *Ecology Letters*, *4*(3), 277–287. doi:10.1046/j.1461-0248.2001.00218.x

Budischak, S. A., O'Neal, D., Jolles, A. E., & Ezenwa, V. O. (2018). Differential host responses to parasitism shape divergent fitness costs of infection. *Functional Ecology*, *32*(2), 324–333. doi:10.1111/1365-2435.12951

Burpee, L. L., & Bouton, J. H. (1993). Effect of eradication of the endophyte Acremonium coenophialum on epidemics of Rhizoctonia blight in tall fescue. *Plant Disease*.

Burpee, L., & Martln, B. (1992). Biology of Rhizoctonia Species Associated with Turfgrasses. *Plant Disease*, *76*(2), 112–117.

Busby, P. E., Peay, K. G., & Newcombe, G. (2016a). Common foliar fungi of Populus trichocarpa modify Melampsora rust disease severity. *New Phytologist*, *209*(4), 1681–1692. doi:10.1111/nph.13742

Busby, P. E., Peay, K. G., & Newcombe, G. (2016b). Common foliar fungi of Populus trichocarpa modify Melampsora rust disease severity. *New Phytologist*, *209*(4), 1681–1692. doi:10.1111/nph.13742

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581.

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Kauff, F., … May, G. (2017). T-BAS : Tree-Based Alignment Selector toolkit for phylogenetic-based placement , alignment downloads and metadata visualization : an example with the Pezizomycotina tree of life, *33*(January 2018), 1160–1168. doi:10.1093/bioinformatics/btw808

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Magain, N., U'Ren, J.M. and F. Lutzoni. 2019. T-BAS version 2.1: Tree-Based Alignment Selector toolkit for evolutionary placement and viewing of alignments and metadata on curated and custom reference trees. Microbiology Resource Announcements In Review

Cattadori, I. M., Boag, B., & Hudson, P. J. (2008). Parasite co-infection and interaction as drivers of host heterogeneity. *International Journal for Parasitology*, *38*(3), 371–380. doi:https://doi.org/10.1016/j.ijpara.2007.08.004

Chaneton, E. J., & Bonsall, M. B. (2000). Enemy-Mediated Apparent Competition: Empirical Patterns and the Evidence. *Oikos*, *88*(2), 380–394.

Choi, Y.-J., Aliota, M. T., Mayhew, G. F., Erickson, S. M., & Christensen, B. M. (2014). Dual RNA-seq of Parasite and Host Reveals Gene Expression Dynamics during Filarial Worm–Mosquito Interactions. *PLOS Neglected Tropical Diseases*, *8*(5), e2905.

Christian, N., Whitaker, B., & Clay, K. (2015). Microbiomes: unifying animal and plant systems through the lens of community ecology theory. *Frontiers in Microbiology, 6,* 869.

Clay, K., Cheplick, G. P., & Marks, S. (1989). Impact of the Fungus Balansia henningsiana on Panicum agrostoides: Frequency of Infection, Plant Growth and Reproduction, and Resistance to Pests. *Oecologia*, *80*(3), 374–380.

Clay, P. A., Dhir, K., Rudolf, V. H. W., & Duffy, M. A. (2018). Within-Host Priority Effects Systematically Alter Pathogen Coexistence. *The American Naturalist*, *193*(2), 187-199. doi:10.1086/701126

Commission, I. O. of E. A. A. H. S. (2006). *Manual of diagnostic tests for aquatic animals*. Office International des Epizooties.

Consortium, T. H. M. P., Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., … White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*, 207.

Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M., & Relman, D. A. (2012). The Application of Ecological Theory Toward an Understanding of the Human Microbiome. *Science*, *336*(6086), 1255 LP-1262. doi:10.1126/science.1224203

Cox FE. (2001). Concomitant infections, parasites and immune responses. *Parasitology*, *122*, S23–S38.

Darling, E. S., & Côté, I. M. (2008). Quantifying the evidence for ecological synergies. *Ecology Letters*, *11*(12), 1278–1286. doi:10.1111/j.1461-0248.2008.01243.x

de Mendiburu, F., & de Mendiburu, M. F. (2015). Package 'agricolae.' *R Package, Version*, 1–2.

Diuk-Wasser, M. A., Vannier, E., & Krause, P. J. (2016). Coinfection by Ixodes Tick-Borne Pathogens: Ecological, Epidemiological, and Clinical Consequences. *Trends in Parasitology*, *32*(1), 30–42. doi:10.1016/j.pt.2015.09.008

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 1–7. doi: 10.1093/bioinformatics/bts635

Ekblom, R., & Galindo, J. (2010). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1–15. doi:10.1038/hdy.2010.152

Ellison, A. R., DiRenzo, G. V, McDonald, C. A., Lips, K. R., & Zamudio, K. R. (2017). First in Vivo Batrachochytrium dendrobatidis Transcriptomes Reveal Mechanisms of Host Exploitation, Host-Specific Gene Expression, and Expressed Genotype Shifts. *G3: Genes|Genomes|Genetics*, *7*(1), 269 LP-278.

Eren, A. M., Borisy, G. G., Huse, S. M., & Mark Welch, J. L. (2014). Oligotyping analysis of the human oral microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), E2875–E2884. doi:10.1073/pnas.1409644111

Ezenwa, V. O., & Jolles, A. E. (2015). Opposite effects of anthelmintic treatment on microbial infection at individual versus population scales. *Science*, *347*(6218), 175 LP-177. doi:10.1126/science.1261714

Fisher, M. C., Henk, D. a, Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, *484*(7393), 186–94. doi:10.1038/nature10947

Fukami, T. (2015). Historical Contingency in Community Assembly : Integrating Niches , Species Pools , and Priority Effects. doi:10.1146/annurev-ecolsys-110411-160340

Gause, G. F., & Witt, A. A. (1935). Behavior of Mixed Populations and the Problem of Natural Selection. *The American Naturalist*, *69*(725), 596–609.

Gerardo, N. M., & Parker, B. J. (2014). Mechanisms of symbiont-conferred protection against natural enemies: An ecological and evolutionary framework. *Current Opinion in Insect Science*, *4*(1), 8–14. doi:10.1016/j.cois.2014.08.002

Gibson, A. K., Raverty, S., Lambourn, D. M., Huggins, J., Magargal, S. L., & Grigg, M. E. (2011). Polyparasitism Is Associated with Increased Disease Severity in Toxoplasma gondii-Infected Marine Sentinel Species. *PLOS Neglected Tropical Diseases*, *5*(5), e1142.

Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology*, *43*, 205–27. doi:10.1146/annurev.phyto.43.040204.135923

Gónzalez, D., Rodriguez-Carres, M., Boekhout, T., Stalpers, J., Kuramae, E. E., Nakatani, A. K., … Cubeta, M. A. (2016). Phylogenetic relationships of Rhizoctonia fungi within the Cantharellales. *Fungal Biology*, *120*(4), 603–619. doi:https://doi.org/10.1016/j.funbio.2016.01.012

Gorsich, E. E., Etienne, R. S., Medlock, J., Beechler, B. R., Spaan, J. M., Spaan, R. S., … Jolles, A. E. (2018). Opposite outcomes of coinfection at individual and population scales. *Proceedings of the National Academy of Sciences*, *115*(29), 7545 LP-7550. doi:10.1073/pnas.1801095115

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech*, *29*(7), 644–652.

Graham, A. L. (2008). Ecological rules governing helminth-microparasite coinfection. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(2), 566–570. doi:10.1073/pnas.0707221105

Griebel, T., Zacher, B., Ribeca, P., Raineri, E., Lacroix, V., Guigó, R., & Sammeth, M. (2012). Modelling and simulating generic RNA-Seq experiments with the flux simulator. *Nucleic*

Acids Research , *40*(20), 10073–10083. doi:10.1093/nar/gks666

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols*, *8*(8), 10.1038/nprot.2013.084.

Haine, E. R. (2008). Symbiont-mediated protection. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1633), 353–361. doi:10.1098/rspb.2007.1211

Hairston, N. G., Smith, F. E., & Slobodkin, L. B. (1960). Community Structure , Population Control , and Competition. *The American Naturalist*, *94*(879), 421–425.

Halliday, F. W., Heckman, R. W., Wilfahrt, P. A., & Mitchell, C. E. (2017). A multivariate test of disease risk reveals conditions leading to disease amplification. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1865), 20171340. doi:10.1098/rspb.2017.1340

Halliday, F. W., Heckman, R. W., Wilfahrt, P. A., & Mitchell, C. E. (2019). Past is prologue: host community assembly and the risk of infectious disease over time. *Ecology Letters*, *22*(1), 138–148. doi:10.1111/ele.13176

Halliday, F. W., Umbanhowar, J., & Mitchell, C. E. (2017). Interactions among symbionts operate across scales to influence parasite epidemics. *Ecology Letters*, *20*, 1285–1294. doi:10.1111/ele.12825

Halliday, F. W., Umbanhowar, J., & Mitchell, C. E. (2018). A host immune hormone modifies parasite species interactions and epidemics: insights from a field manipulation. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1890), 20182075. doi:10.1098/rspb.2018.2075

Hayden, K. J., Garbelotto, M., Knaus, B. J., Cronn, R. C., Rai, H., & Wright, J. W. (2014). Dual RNA-seq of the plant pathogen Phytophthora ramorum and its tanoak host. *Tree Genetics & Genomes*, *10*(3), 489–502. doi:10.1007/s11295-014-0698-0

Hayes, K. S., Bancroft, A. J., Goldrick, M., Portsmouth, C., Roberts, I. S., & Grencis, R. K. (2010). Exploitation of the Intestinal Microflora by the Parasitic Nematode Trichuris muris; *Science*, *328*(5984), 1391 LP-1394. doi:10.1126/science.1187703

Higgins, K. L., Arnold, a E., Miadlikowska, J., Sarvate, S. D., & Lutzoni, F. (2007). Phylogenetic relationships, host affinity, and geographic structure of boreal and arctic endophytes from three major plant lineages. *Molecular Phylogenetics and Evolution*, *42*(2), 543–55. doi:10.1016/j.ympev.2006.07.012

Hill, M. O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, *54*(2), 427–432. doi:10.2307/1934352

Holling, C. S. (1959). The Components of Predation as Revealed by a Study of Small. Mammal

Predation of the European Pine Sawfly. *The Canadian Entomologist*, *XCI*(5), 234–261.

Hood, M. E. (2003). Dynamics of Multiple Infection and Within-Host Competition by the Anther-Smut Pathogen. *The American Naturalist*, *162*(1), 122–133. doi:10.1086/375539

Hopkins, S. R., Wojdak, J. M., & Belden, L. K. (2017). Defensive Symbionts Mediate Host-Parasite Interactions at Multiple Scales. *Trends in Parasitology*, *33*(1), 53–64. doi:10.1016/j.pt.2016.10.003

Jani, A. J., & Briggs, C. J. (2014). The pathogen Batrachochytrium dendrobatidis disturbs the frog skin microbiome during a natural epidemic and experimental infection. *Proceedings of the National Academy of Sciences* , *111*(47), E5049–E5058. doi:10.1073/pnas.1412752111

Jiang, J.-H., Lee, Y.-I., Cubeta, M. A., & Chen, L.-C. (2015). Characterization and colonization of endomycorrhizal Rhizoctonia fungi in the medicinal herb Anoectochilus formosanus (Orchidaceae). *Mycorrhiza*, *25*(6), 431–445. doi:10.1007/s00572-014-0616-1

Johnson, P. T. J., de Roode, J. C., & Fenton, A. (2015). Why infectious disease research needs community ecology. *Science*, *349*(6252), 1259504-1-1259504–9. doi:10.1126/science.1259504

Johnson, P. T. J., & Hoverman, J. T. (2012). Parasite diversity and coinfection determine pathogen infection success and host fitness. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(23), 9006–11. doi:10.1073/pnas.1201790109

Jost, L. (2006). Entropy and diversity. *Oikos*, *113*(2), 363–375. doi:10.1111/j.2006.0030-1299.14714.x

Kauppinen, M., Saikkonen, K., Helander, M., Pirttilä, A. M., & Wäli, P. R. (2016). Epichloë grass endophytes in sustainable agriculture. *Nature Plants*, *2*(2), 15224. doi:10.1038/nplants.2015.224

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T., & Minami, E. (2012). Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PloS One*, *7*(11), e49423. doi:10.1371/journal.pone.0049423

Koch, H., & Schmid-Hempel, P. (2011). Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proceedings of the National Academy of Sciences*, *108*(48), 19288 LP-19292. doi:10.1073/pnas.1110474108

Krauss, J., Härri, S.A., Bush, L., Power, S.A., Muller, C.B. (2007). Fungal grass endophytes, grass cultivars, nitrogen deposition and the associations with colonizing insects. Proceedings of the 6th International Symposium on Fungal Endophytes of Grasses, pp. 53 57.

Kreisinger, J., Bastien, G., Hauffe, H. C., Marchesi, J., & Perkins, S. E. (2015). Interactions

between multiple helminths and the gut microbiota in wild rodents. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1675), 20140295. doi:10.1098/rstb.2014.0295

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. doi:10.1038/nmeth.1923

Lee, K., Pan, J. J., & May, G. (2009). Endophytic Fusarium verticillioides reduces disease severity caused by Ustilago maydis on maize. *FEMS Microbiology Letters*, *299*(1), 31–37. doi:10.1111/j.1574-6968.2009.01719.x

Lee, S. C., Tang, M. S., Lim, Y. A. L., Choy, S. H., Kurtz, Z. D., Cox, L. M., … Loke, P. (2014). Helminth colonization is associated with increased diversity of the gut microbiota. *PLoS Neglected Tropical Diseases*, *8*(5), e2880–e2880. doi:10.1371/journal.pntd.0002880

Leggett, H. C., Cornwallis, C. K., Buckling, A., & West, S. A. (2017). Growth rate, transmission mode and virulence in human pathogens. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1719). doi:10.1098/rstb.2016.0094

Lenth, R. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R Package Version*, *1*(1).

Leung, J. M., Budischak, S. A., Chung The, H., Hansen, C., Bowcutt, R., Neill, R., … Graham, A. L. (2018). Rapid environmental effects on gut nematode susceptibility in rewilded mice. *PLOS Biology*, *16*(3), e2004108.

Li, C., Gao, J., & Nan, Z. (2007). Interactions of Neotyphodium gansuense, Achnatherum inebrians, and plant-pathogenic fungi. *Mycological Research*, *111*(10), 1220–1227. doi:https://doi.org/10.1016/j.mycres.2007.08.012

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* , *26*(5), 589–595. doi:10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi:10.1093/bioinformatics/btp352

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, *41*(10), e108–e108. doi:10.1093/nar/gkt214

Libertucci, J., & Young, V. B. (2019). The role of the microbiota in infectious diseases. *Nature Microbiology*, *4*(1), 35–45. doi:10.1038/s41564-018-0278-4

Lohr, J. N., Yin, M., & Wolinska, J. (2010). Prior residency does not always pay off – co-infections in Daphnia. *Parasitology*, *137*(10), 1493–1500. doi:DOI:

10.1017/S0031182010000296

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. doi:10.1186/s13059-014-0550-8

Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., … Dangl, J. L. (2012). Defining the core Arabidopsis thaliana root microbiome. *Nature*, *488*(7409), 86–90.

Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, *10*(10), 999–1002. doi:10.1038/nmeth.2634

Malinowski, D. P., & Belesky, D. P. (2000). Adaptations of Endophyte-Infected Cool-Season Grasses to Environmental Stresses: Mechanisms of Drought and Mineral Stress Tolerance. *Crop Science*, *40*, 923–940. doi:10.2135/cropsci2000.404923x

Marchetto, K. M., & Power, A. G. (2017). Coinfection Timing Drives Host Population Dynamics through Changes in Virulence. *The American Naturalist*, *191*(2), 173–183. doi:10.1086/695316

Marchetto, K. M., & Power, A. G. (2018). Context-dependent interactions between pathogens and a mutualist affect pathogen fitness and mutualist benefits to hosts. *Ecology*, *99*(12), 2833–2843. doi:10.1002/ecy.2531

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal; Vol 17, No 1: Next Generation Sequencing Data Analysis DO - 10.14806/Ej.17.1.200*.

May, G., & Nelson, P. (2014). Defensive mutualisms: do microbial interactions within hosts drive the evolution of defensive traits? *Functional Ecology*, *28*(2), 356–363. doi:10.1111/1365-2435.12166

Mitchell, C. E., Tilman, D., & Groth, J. V. (2002). Effects of Grassland Plant Species Diversity, Abundance, and Composition on Foliar Fungal Disease. *Ecology*, *83*(6), 1713–1726.

Mitchell, C. E., Tilman, D., & Groth, J. V. (2003). Effects of elevated CO2, nitrogen deposition, and decreased species diversity on foliar fungal plant disease. *Global Change Biology*, *9*(3), 438–451. doi:10.1046/j.1365-2486.2003.00602.x

Mordecai, E. A., Gross, K., & Mitchell, C. E. (2015). Within-Host Niche Differences and Fitness Trade-offs Promote Coexistence of Plant Viruses. *The American Naturalist*, *187*(1), E13–E26. doi:10.1086/684114

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and

quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth*, *5*(7), 621–628.

Mosca, A., Leclerc, M., & Hugot, J. P. (2016). Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem?. *Frontiers in Microbiology* .

O'Keeffe, K. R., Carbone, I., Jones, C. D., & Mitchell, C. E. (2017). Plastic potential: how the phenotypes and adaptations of pathogens are influenced by microbial interactions within plants. *Current Opinion in Plant Biology*, *38*, 78–83. doi:10.1016/j.pbi.2017.04.014

O'Keeffe, K. R., & Jones, C. D. (2019). Challenges and solutions for analysing dual RNA-seq data for non-model host–pathogen systems. *Methods in Ecology and Evolution*, *10*(3), 401–414. doi:10.1111/2041-210X.13135

Oliver, K. M., Moran, N. A., & Hunter, M. S. (2005). Variation in resistance to parasitism in aphids is due to symbionts not host genotype. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(36), 12795 LP-12800. doi:10.1073/pnas.0506131102

Panaccione, D. G., Beaulieu, W. T., & Cook, D. (2014). Bioactive alkaloids in vertically transmitted fungal endophytes. *Functional Ecology*, *28*(2), 299–314. doi:10.1111/1365-2435.12076

Pańka, D., West, C. P., Guerber, C. a., & Richardson, M. D. (2013). Susceptibility of tall fescue to Rhizoctonia zeae infection as affected by endophyte symbiosis. *Annals of Applied Biology*, *163*(2), 257–268. doi:10.1111/aab.12051

Pathirana, N. U. K., Meegaskumbura, M., & Rajakaruna, R. S. (2019). Infection sequence alters disease severity—Effects of the sequential exposure of two larval trematodes to *Polypedates cruciger* tadpoles. *Ecology and Evolution*, (December 2018), ece3.5180. doi:10.1002/ece3.5180

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2013). nlme: Linear and nonlinear mixed effects models. *R Package Version*, *3*(1), 111.

Pinna, R. A., Silva-dos-Santos, D., Perce-da-Silva, D. S., Oliveira-Ferreira, J., Villa-Verde, D. M. S., De Luca, P. M., & Banic, D. M. (2016). Malaria-Cutaneous Leishmaniasis Co-infection: Influence on Disease Outcomes and Immune Response . *Frontiers in Microbiology (7),* 962.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* , *26*(6), 841–842. doi:10.1093/bioinformatics/btq033

Rhind, N., Chen, Z., Yassour, M., Thompson, D. A., Haas, B. J., Habib, N., … Nusbaum, C. (2011). Comparative Functional Genomics of the Fission Yeasts. *Science*, *332*(6032), 930 LP-936.

Rosen, M. J., Callahan, B. J., Fisher, D. S., & Holmes, S. P. (2012). Denoising PCR-amplified metagenome data. *BMC Bioinformatics*, *13*, 283. doi:10.1186/1471-2105-13-283

Saikkonen, K., Faeth, S. H., Helander, M., & Sullivan, T. J. (1998). Fungal Endophytes: A Continuum of Interactions with Host Plants. *Annual Review of Ecology and Systematics*, *29*(1), 319–343. doi:10.1146/annurev.ecolsys.29.1.319

Saikkonen, K., Gundel, P. E., & Helander, M. (2013). Chemical Ecology Mediated by Fungal Endophytes in Grasses. *Journal of Chemical Ecology*, *39*(7), 962–968. doi:https://doi.org/10.1007/s10886-013-0310-3

Saikkonen, K., Helander, M., Faeth, S. H., Schulthess, F., & Wilson, D. (1999). Endophyte-grass-herbivore interactions: The case of Neotyphodium endophytes in Arizona fescue populations. *Oecologia*, *121*(3), 411–420. doi:10.1007/s004420050946

Saikkonen, K., Young, C. A., Helander, M., & Schardl, C. L. (2016). Endophytic Epichloë species and their grass hosts: from evolution to applications. *Plant Molecular Biology*, *90*(6), 665–675. doi:10.1007/s11103-015-0399-6

Santhanam, R., Luu, V. T., Weinhold, A., Goldberg, J., Oh, Y., & Baldwin, I. T. (2015). Native root-associated bacteria rescue a plant from a sudden-wilt disease that emerged during continuous cropping. *Proceedings of the National Academy of Sciences*, *112*(36), E5013 LP-E5020. doi:10.1073/pnas.1505765112

Schloss, P. D., & Handelsman, J. (2005). Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. *Applied and Environmental Microbiology*, *71*(3), 1501–1506. doi:10.1128/AEM.71.3.1501

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., … Schindel, D. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16), 6241–6246. doi:10.1073/pnas.1117018109

Sedlazeck, F. J., Rescheneder, P., & von Haeseler, A. (2013). NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* , *29*(21), 2790–2791. doi:10.1093/bioinformatics/btt468

Selosse, M.-A., Bessis, A., & Pozo, M. J. (2014). Microbial priming of plant and animal immunity: symbionts as developmental signals. *Trends in Microbiology*, *22*(11), 607–613. doi:10.1016/j.tim.2014.07.003

Sih, A., Englund, G., & Wooster, D. (1998). Emergent impacts of multiple predators on prey. *Trends in Ecology & Evolution*, *13*(9), 350–355. doi:10.1016/S0169-5347(98)01437-2

Simko, I., & Piepho, H.-P. (2011). The Area Under the Disease Progress Stairs: Calculation, Advantage, and Application. *Phytopathology*, *102*(4), 381–389. doi:10.1094/PHYTO-07-

11-0216

Smith, D. P., & Peay, K. G. (2014). Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS ONE*, *9*(2). doi:10.1371/journal.pone.0090234

Spear, E. R., & Mordecai, E. A. (2018). Foliar pathogens are unlikely to stabilize coexistence of competing species in a California grassland. *Ecology*, *99*(10), 2250–2259. doi:10.1002/ecy.2427

Sun, Y., Cai, Y., Liu, L., Yu, F., Farrell, M. L., Mckendree, W., & Farmerie, W. (2009). ESPRIT: Estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Research*, *37*(10). doi:10.1093/nar/gkp285

Susi, H., Barrès, B., Vale, P. F., & Laine, A.-L. (2015). Co-infection alters population dynamics of infectious disease. *Nature Communications*, *6*(Viikinkaari 1), 5975. doi:10.1038/ncomms6975

Teixeira, P. J. P. L., Thomazella, D. P. de T., Reis, O., do Prado, P. F. V., do Rio, M. C. S., Fiorin, G. L., … Pereira, G. A. G. (2014). High-Resolution Transcript Profiling of the Atypical Biotrophic Interaction between Theobroma cacao and the Fungal Pathogen Moniliophthora perniciosa. *The Plant Cell*, *26*(11), 4245–4269. doi:10.1105/tpc.114.130807

Telfer, S., Lambin, X., Birtles, R., Beldomenico, P., Burthe, S., Paterson, S., & Begon, M. (2010). Species Interactions in a Parasite Community Drive Infection Risk in a Wildlife Population. *Science (New York, N.Y.)*, *330*(6001), 243–246. doi:10.1126/science.1190333

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling Survivl Data: Extending the Cox Model*. New York: Springer.

Tian, P., Nan, Z., Li, C., & Spangenberg, G. (2008). Effect of the endophyte Neotyphodium lolii on susceptibility and host physiological response of perennial ryegrass to fungal pathogens. *European Journal of Plant Pathology*, *v. 122*(4), 593-602–2008 v.122 no.4. doi:10.1007/s10658-008-9329-7

Tollenaere, C., Susi, H., & Laine, A. (2016). Evolutionary and Epidemiological Implications of Multiple Infection in Plants. *Trends in Plant Science*, *21*(1), 80–90. doi:10.1016/j.tplants.2015.10.014

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* , *25*(9), 1105–1111. doi:10.1093/bioinformatics/btp120

Trinh, P., Zaneveld, J. R., Safranek, S., & Rabinowitz, P. M. (2018). One Health Relationships Between Human, Animal, and Environmental Microbiomes: A Mini-Review. *Frontiers in Public Health*, *6*, 235. doi:10.3389/fpubh.2018.00235

Turner, T. R., James, E. K., & Poole, P. S. (2013). The plant microbiome. *Genome Biology*, *14*(6), 209. doi:10.1186/gb-2013-14-6-209

Viney, M. E., & Graham, A. L. (2013). Chapter Five - Patterns and Processes in Parasite Co-Infection. In D. B. T.-A. in P. Rollinson (Ed.) (Vol. 82, pp. 321–369). Academic Press. doi:https://doi.org/10.1016/B978-0-12-407706-5.00005-8

Wäli, P. R., Helander, M., Nissinen, O., & Saikkonen, K. (2006). Susceptibility of endophyte-infected grasses to winter pathogens (snow molds). *Canadian Journal of Botany*, *84*(7), 1043–1051.

Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., … Liu, J. (2010). MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Research*, *38*(18), e178–e178.

Westermann, A. J., Barquist, L., & Vogel, J. (2017). Resolving host–pathogen interactions by dual RNA-seq. *PLOS Pathogens*, *13*(2), e1006033.

Westermann, A. J., Förstner, K. U., Amman, F., Barquist, L., Chao, Y., Schulte, L. N., … Vogel, J. (2016). Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. *Nature*, *529*, 496.

Westermann, A. J., Gorski, S. A., & Vogel, J. (2012). Dual RNA-seq of pathogen and host. *Nat Rev Microbiol*, *10*. doi:10.1038/nrmicro2852

White, T. J. (1990). Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In *A Guide to Methods and Applications* (pp. 315–322).

Wintermantel, W. M., Cortez, A. A., Anchieta, A. G., Gulati-Sakhuja, A., & Hladky, L. L. (2008). Co-Infection by Two Criniviruses Alters Accumulation of Each Virus in a Host-Specific Manner and Influences Efficiency of Virus Transmission. *Phytopathology*, *98*(12), 1340–1345. doi:10.1094/PHYTO-98-12-1340

Wootton, J. T. (1994). The Nature and Consequences of Indirect Effects in Ecological Communities. *Annual Review of Ecology and Systematics*, *25*(1), 443–466. doi:10.1146/annurev.es.25.110194.002303

Wu, S. B., Stanley, D., Rodgers, N., Swick, R. A., & Moore, R. J. (2014). Two necrotic enteritis predisposing factors, dietary fishmeal and Eimeria infection, induce large changes in the caecal microbiota of broiler chickens. *Veterinary Microbiology*, *169*(3–4), 188–197. doi:10.1016/j.vetmic.2014.01.007

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W, editors. New York, NY: Spring Science and Business Media.